J. Appl. Prob. **55,** 1186–1202 (2018)
*doi:*10.1017/jpr.2018.78

COMPARISON OF HIT-AND-RUN, SLICE SAMPLER AND RANDOM WALK METROPOLIS

DANIEL RUDOLF,* *Universität Göttingen*
MARIO ULLRICH,** *Universität Linz*

Abstract

Different Markov chains can be used for approximate sampling of a distribution given by an unnormalized density function with respect to the Lebesgue measure. The hit-and-run, (hybrid) slice sampler, and random walk Metropolis algorithm are popular tools to simulate such Markov chains. We develop a general approach to compare the efficiency of these sampling procedures by the use of a partial ordering of their Markov operators, the covariance ordering. In particular, we show that the hit-and-run and the simple slice sampler are more efficient than a hybrid slice sampler based on hit-and-run, which, itself, is more efficient than a (lazy) random walk Metropolis algorithm.

Keywords: Hit-and-run; slice sampler; random walk Metropolis; covariance ordering

2010 Mathematics Subject Classification: Primary 60J22
Secondary 65C05; 60J05

1. Introduction

In many scenarios of Bayesian statistics, statistical physics, and other branches of applied sciences (see [5], [17], and [32]), it is of interest to sample on $\mathbb{R}^d$ with respect to a distribution π. In particular, we assume that π is given by an unnormalized density. More precisely, let $K \subseteq \mathbb{R}^d$ be an open, measurable set and $\rho\colon K \to (0,\infty)$ be a positive, bounded and with respect to the Lebesgue measure an almost everywhere continuous function with $\int_K \rho(x)\,\mathrm{d}x \in (0,\infty)$. Then, define the probability measure π through ρ by

$$\pi(A) = \frac{\int_A \rho(x)\,\mathrm{d}x}{\int_K \rho(x)\,\mathrm{d}x} \quad \text{for all measurable } A \subseteq K.$$

Arguably the most successful approach to approximate π is the construction of a suitable Markov chain. The hit-and-run algorithm, the random walk Metropolis, the simple slice sampler, and hybrid slice sampler provide such construction methods. The crucial question is which one of these algorithm should we use?

This question is, of course, related to the speed of convergence of the Markov chain sampling and any answer depends very much on the imposed assumptions. In general, it is difficult to derive explicit estimates of this speed of convergence. But it might be possible to prove that one algorithm is better than another. This motivates the idea of the comparison of Markov chains.

The first comparison result of Markov chains is due to Peskun [29]. There, a partial ordering on finite state spaces was invented, where one transition kernel has higher order than another

Received 4 July 2017; revision received 17 August 2018.
* Postal address: Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, Universität Göttingen, Goldschmidstraße 7, 37077 Göttingen, Germany. Email address: daniel.rudolf@uni-goettingen.de
** Postal address: Universität Linz, Altenberger Straße 69, 4040 Linz, Austria. Email address: mario.ullrich@jku.at

one if the former dominates the latter off the diagonal. This order was later extended by Tierney [39] to general state spaces. However, for the Markov chains we have in mind it does not seem possible to use this off-diagonal ordering. We consider a partial ordering on the set of linear operators; see [14, p. 470]. In the context of Markov chains, this ordering is called covariance ordering; see [22] and [23]. Let $L_2(\pi)$ be the Hilbert space of functions with finite stationary variance and assume that $P_1, P_2 \colon L_2(\pi) \to L_2(\pi)$ are two self-adjoint linear operators. Then we say that $P_1 \leq P_2$ if and only if

$$\langle P_1 f, f\rangle_\pi \geq \langle P_2 f, f\rangle_\pi, \qquad f \in L_2(\pi).$$

Here, the inner product of $L_2(\pi)$ is

$$\langle f_1, f_2\rangle_\pi = \int_K f_1(x) f_2(x)\,\mathrm{d}\pi(x), \qquad f_1, f_2 \in L_2(\pi).$$

The operator P_1 is called positive if $P_1 \geq 0$, i.e. $\langle P_1 f, f\rangle_\pi \geq 0$ for any $f \in L_2(\pi)$. Reversible transition kernels of given Markov chains induce self-adjoint Markov operators and we can compare these operators.

Let P_1, P_2 be two such Markov operators and assume that $0 \leq P_1 \leq P_2$. There are a number of consequences for the corresponding Markov chains: for example, the spectral gap of P_2 is smaller than that of P_1. The spectral gap of a Markov chain is a quantity which is closely related to the speed of convergence to π; see [3] and [35]. Another example concerns the stationary asymptotic variance of sample averages. For $i = 1, 2$, let $S_n^{(i)}(f) = (1/n)\sum_{k=1}^n f(X_k^{(i)})$ with Markov chain $(X_k^{(i)})_{k\in\mathbb{N}}$ starting in stationarity corresponding to P_i and an arbitrary function $f \in L_2(\pi)$. These sample averages yield approximations of the mean $Ef = \int_K f\,\mathrm{d}\pi$ and in [22] and [39] it was observed that if $P_1 \leq P_2$ then

$$V(f, P_2) \leq V(f, P_1).$$

Here $V(f, P_i) := \lim_{n\to\infty} n\mathbb{E}|S_n^{(i)}(f) - Ef|^2$ is the stationary asymptotic variance of $S_n^{(i)}(f)$. In other words, the Markov chain of P_2 is (asymptotically) more efficient than that of P_1. For more details of the implications of $P_1 \leq P_2$, we refer the reader to Section 2.1.1.

Our aim in this paper is to compare the hit-and-run algorithm, the (lazy) random walk Metropolis, the simple slice sampler, and a hybrid slice sampler based on hit-and-run according to this partial ordering. To do so we develop a systematic approach for the comparison of Markov chains which can be written by a suitable two-step procedure.

The intuition behind our approach is the following. It is a simple and well-known observation that, if the self-adjoint and positive Markov operators P_1 and P_2 satisfy $P_1 P_2 = P_2$, then $P_1 \leq P_2$. However, due to the quite restrictive assumptions, this technique can be used only in a small number of cases. A useful generalization of this technique, which was invented in a special case in [41], can be used whenever we have the representation $P_i = R\widetilde{P}_i R^*$ for certain operators R and $\widetilde{P}_i$, $i = 1, 2$. In this case, and if $\widetilde{P}_1\widetilde{P}_2 = \widetilde{P}_2$, one can also conclude that $P_1 \leq P_2$; see Lemma 1. Our comparison inequalities therefore follow once we have established such representations for the Markov chains under consideration.

1.1. The algorithms

Let us briefly explain the algorithms. Roughly, a transition of the hit-and-run algorithm works as follows. Choose randomly a line through the current state and sample according to π restricted to this line. Thus, instead of sampling with respect to π on $K \subseteq \mathbb{R}^d$, hit-and-run

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

uses sampling with respect to π only on one-dimensional lines through the state space, which is feasible in many cases. For example, if ρ is a log-concave density then the restriction of π to a line has also a log-concave density and one can use different acceptance/rejection methods to sample such a distribution on the line efficiently. For details we refer the reader to [6, Section 2.4.2]; see also [19] and [36].

In contrast, the simple slice sampler chooses a suitable d-dimensional set, a super-level set of ρ, depending on the current state and then samples the next state of the Markov chain uniformly distributed on this super-level set. Sampling of the uniform distribution on a d-dimensional set is often not efficiently implementable. This is why this Markov chain is mostly of theoretical interest.

The hybrid slice sampler we are interested in overcomes this problem by replacing the uniform sampling by one step of a hit-and-run algorithm on the super-level set. First, choose a line through the current state uniformly at random and then generate the next state uniformly distributed on the intersection of that line with the super-level set. We call this procedure a hybrid slice sampler based on hit-and-run. Intuitively, it is clear that the simple slice sampler is better than that hybrid one. The intuition for the comparison of the hit-and-run algorithm and the hybrid slice sampler based on hit-and-run is not so obvious. Observe that this particular hybrid sampler can also be interpreted as choosing first a line and then performing a simple slice sampling step according the distribution of π restricted to that line. This observation leads us to the fact that the hit-and-run algorithm is better.

Finally, we consider the random walk Metropolis algorithm. Assume that we have a proposal density q on $\mathbb{R}^d$ and let $x \in K$ be the current state. A transition works as follows: generate $z \in \mathbb{R}^d$ according the distribution determined by q and accept $x + z$ as the next state with probability $\min\{1, \rho(x+z)/\rho(x)\}$ if $x + z \in K$. Otherwise, stay at x. It is well known (see [10] and [37]) that the random walk Metropolis algorithm can be interpreted as a certain (hybrid) slice sampling procedure which runs a random walk according to q on the super-level set with uniform limit distribution. We want to compare the random walk Metropolis and the hybrid slice sampler based on a hit-and-run. Thus, the question is whether the uniform hit-and-run step is better than the random walk step on the super-level set. It turns out that this is indeed the case.

1.2. Main results

Now let us formulate the main results. To compare the above Markov chains we develop, in Section 3, a general approach which might be of interest in its own right. There, in Lemma 1, conditions for two suitably defined Markov operators P_1, P_2 are stated, which imply that $P_1 \leq P_2$. This lemma is the main ingredient for the comparison argument. Its application leads us to Theorem 1. For the Markov operators M, U, H, S of the (lazy) random walk Metropolis with rotational invariant proposal q, the hybrid slice sampler based on hit-and-run, the hit-and-run algorithm, and the simple slice sampler, respectively, we have

$$M \leq U \leq H, \qquad M \leq U \leq S.$$

Thus, the random walk Metropolis is less efficient than the hit-and-run algorithm and simple slice sampler. The hybrid slice sampler based on hit-and-run we propose lies somewhere in between in terms of efficiency.

1.3. Outline

The paper is organized as follows. In the next section we introduce our notation, comment on the partial ordering, present consequences for the Markov chains, and state the algorithms

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

we study in detail. In Section 3 we invent a new approach of how to compare Markov chains with a specific structure. Section 4 contains the application of the former developed comparison arguments. Finally, in Section 5, we provide some concluding remarks and discuss open problems.

2. Preliminaries

Let $L_2(\pi)$ be the Hilbert space of functions $f\colon K \to \mathbb{R}$ with finite norm $\|f\|_\pi = \langle f, f\rangle_\pi^{1/2}$, where the inner product of $f_1, f_2 \in L_2(\pi)$ is

$$\langle f_1, f_2\rangle_\pi = \int_K f_1(x) f_2(x)\,\mathrm{d}\pi(x).$$

Let P be a transition kernel on K which is reversible with respect to π and let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel P, i.e.

$$P(x, A) = \mathbb{P}(X_{k+1} \in A \mid X_k = x)$$

almost surely for all $k \in \mathbb{N}$ and $A \subseteq K$. The corresponding Markov operator, also denoted by P, is

$$Pf(x) = \int_K f(y) P(x, \mathrm{d}y) \quad \text{for } f \in L_2(\pi).$$

Obviously, $P(x, A) = P\,\mathbf{1}_A(x)$ for all $A \subseteq K$, where $\mathbf{1}_A$ denotes the indicator function of A. Note that, by reversibility, $P\colon L_2(\pi) \to L_2(\pi)$ is self-adjoint. We say that a (Markov) operator P on $L_2(\pi)$ is *positive* if $\langle Pf, f\rangle_\pi \geq 0$ for all $f \in L_2(\pi)$.

2.1. On the ordering and consequences

With this notation we define on the set of Markov operators the following partial ordering. For Markov operators P_1 and P_2, we write

$$P_1 \leq P_2 \quad \Longleftrightarrow \quad \langle P_1 f, f\rangle_\pi \geq \langle P_2 f, f\rangle_\pi \quad \text{for all } f \in L_2(\pi).$$

In the following, we discuss why the choice of this ordering is particularly meaningful for Markov chains.

2.1.1. *Consequences for the speed of convergence.* There are many ways to measure the speed of convergence of the distribution of a Markov chain towards its stationary distribution. Probably the most desirable quantity is the *total variation distance* of νP^n and π, i.e.

$$\|\nu P^n - \pi\|_{\mathrm{TV}} = \sup_{A\subseteq K} |\nu P^n(A) - \pi(A)|,$$

where $\nu P^n(A) = \int_K P^n(x, A)\,\mathrm{d}\nu(x)$ is the distribution of the Markov chain with transition kernel P and initial distribution ν after n steps. However, estimating the total variation distance is quite delicate and, in practice, it is usually much easier to derive bounds on it by more analytic quantities, such as isoperimetric constants or certain norms of P; see, for example, [16], [18], and [20]. Many of these quantities are defined by

$$c_{\mathcal{M}}(P) = \inf_{f\in\mathcal{M}} \langle (I - P) f, f\rangle_\pi$$

for certain sets of functions $\mathcal{M} \subset L_2(\pi)$, where $If := f$. Hence, in the proof of $P_1 \leq P_2$, it is enough to obtain $c_{\mathcal{M}}(P_1) \leq c_{\mathcal{M}}(P_2)$ for every choice of $\mathcal{M}$. Prominent examples are as follows:

- spectral gap: $\mathcal{M} = \{f\colon \|f\|_\pi = 1, \int_K f\,\mathrm{d}\pi = 0\}$;

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

- conductance: $\mathcal{M} = \{f\colon f = \mathbf{1}_A/\sqrt{\pi(A)}, \pi(A) \in (0, \frac{1}{2}], A \subset K\}$;
- log-Sobolev constant: $\mathcal{M} = \{f\colon \int_K f^2 \log(f^2/\|f\|_\pi^2)\,\mathrm{d}\pi = 1\}$.

There are more quantities of this form, such as the best constant in a Nash inequality, and average and blocking conductance; see, for example, [7] and [26] for details.

In what follows, we will prove $P_1 \le P_2$ for a couple of Markov operators and, hence, that P_2 is 'faster' than P_1 in all the above senses.

2.1.2. *Consequences for sample averages.* The property $P_1 \le P_2$ also has consequences for the worst case mean-square error and the asymptotic stationary variance of Markov chain Monte Carlo methods.

Let $(X_k^{(i)})_{k\in\mathbb{N}}$ be a Markov chain with transition kernel P_i, $i = 1, 2$, and initial distribution π, and define the *Markov chain Monte Carlo method*

$$S_n^{(i)}(f) = \frac{1}{n}\sum_{k=1}^{n} f(X_k^{(i)}), \qquad i = 1, 2,$$

which gives an approximation to $Ef := \int_K f\,\mathrm{d}\pi$ for $f \in L_2(\pi)$. By virtue of spectral theoretic arguments, we can show that

$$\sup_{\|f\|_\pi \le 1} \mathbb{E}|S_n^{(2)}(f) - Ef|^2 \le \sup_{\|f\|_\pi \le 1} \mathbb{E}|S_n^{(1)}(f) - Ef|^2$$

(we use [35, Corollary 3.27 and Lemma 2.12], as well as the fact that the spectral gap of P_1 is smaller than that of P_2). Another interesting consequence can be found in [23, Theorem 6]. There it was proved that $P_1 \le P_2$ *if and only if*

$$V(f, P_2) \le V(f, P_1) \quad \text{for each individual } f \in L_2(\pi),$$

where $V(f, P_i) := \lim_{n\to\infty} n\mathbb{E}|S_n^{(i)}(f) - Ef|^2$ is the asymptotic stationary variance (which can also be considered as the asymptotic mean-square error).

2.2. The algorithms

We present the different algorithms we consider in detail and provide relevant literature.

2.2.1. *Hit-and-run algorithm.* The hit-and-run algorithm proposed by Smith [38] has been well studied in different settings; see [4], [8], [12], [13], [16], [18], and [36].

Informally, it samples at each step on a randomly chosen one-dimensional line with respect to the corresponding conditional distribution. Let S_{d-1} be the Euclidean unit sphere in $\mathbb{R}^d$. For $x \in K$ and $\theta \in \mathsf{S}_{d-1}$, we define

$$L(x, \theta) = \{x + s\theta \in K \mid s \in \mathbb{R}\}$$

as the *chord through x in direction θ*. A transition from x to y of the hit-and-run algorithm works as follows: generate a set $L(x, \theta)$ by choosing θ with the uniform distribution on the sphere and, then, choose $y \in L(x, \theta)$ according to the distribution determined by ρ conditioned on the chord $L(x, \theta)$. We present a transition of hit-and-run in Algorithm 1.

Algorithm 1. (*Hit-and-run.*) Transition from current state x to next state y.

1. Sample $\theta \sim \mathrm{uniform}(\mathsf{S}_{d-1})$.

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

2. Sample $y \sim H_\theta(x, \cdot)$, where

$$H_\theta(x, A) = \frac{\int_{L(x,\theta)} \mathbf{1}_A(z)\rho(z)\,\mathrm{d}z}{\int_{L(x,\theta)} \rho(z)\,\mathrm{d}z}.$$

Note that the integral in H_θ is over a one-dimensional subset of $\mathbb{R}^d$ and the integration is with respect to the one-dimensional Lebesgue measure. For $x \in K$ and $A \subseteq K$, the transition kernel, say H, of the hit-and-run algorithm is determined by

$$H(x, A) = \int_{\mathsf{S}_{d-1}} H_\theta(x, A)\,\mathrm{d}\sigma(\theta),$$

where $\sigma = \text{uniform}(\mathsf{S}_{d-1})$ denotes the uniform distribution on the sphere. It is well known that this transition kernel is reversible with respect to π; see, for example, [4].

An important special case of the hit-and-run algorithm above is when the density is an indicator function, for example, $\rho = \mathbf{1}_K$. In this case, the hit-and-run algorithm is reversible with respect to the uniform distribution on K. Thus, under weak regularity assumptions, the uniform distribution is the (unique) stationary distribution; see [4]. We call this special case uniform hit-and-run. Let us mention that we use the uniform hit-and-run in the next section.

2.2.2. *Simple and hybrid slice sampling.* Slice sampling belongs to the class of auxiliary variable algorithms that are defined by a Markov chain on an extended state space; see [10], [24], [25], [28], [30], [32], and the references therein.

We consider the simple slice sampler and a hybrid slice sampler based on hit-and-run. A single transition of the simple slice sampler is presented in Algorithm 2.

Algorithm 2. (*Simple slice sampler.*) Transition from current state x to next state y.

1. Sample $t \sim \text{uniform}(0, \rho(x))$.
2. Sample $y \sim \text{uniform}(K(t))$, where

$$K(t) = \{x \in K \mid \rho(x) > t\}$$

is the super-level set of ρ determined by t.

The transition kernel, say S, corresponding to Algorithm 2 is

$$S(x, A) = \frac{1}{\rho(x)} \int_0^{\rho(x)} \frac{\text{vol}_d(A \cap K(t))}{\text{vol}_d(K(t))}\,\mathrm{d}t,$$

where vol_d denotes the d-dimensional Lebesgue measure. The simple slice sampler exhibits quite robust convergence properties; see [24] and [30]. However, a crucial drawback is that the second step is difficult to implement. Because of this we consider the following hybrid slice sampler based on hit-and-run. The idea is to replace the second step of the simple slice sampler by a Markov chain transition according to the uniform hit-and-run algorithm in $K(t)$, and we do this in the next algorithm.

Algorithm 3. (*Hybrid slice sampler based on hit-and-run.*) Transition from current state x to next state y.

1. Sample $t \sim \text{uniform}(0, \rho(x))$ and $\theta \sim \text{uniform}(\mathsf{S}_{d-1})$ independently.

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

2. Sample $y \sim \text{uniform}(L_t(x,\theta))$, where

$$L_t(x,\theta) = \{x + r\theta \in K(t) \mid r \in \mathbb{R}\}$$

is the chord through x in direction θ restricted to $K(t)$.

The transition kernel, say U, of this hybrid slice sampler is

$$U(x, A) = \frac{1}{\rho(x)} \int_0^{\rho(x)} \int_{\mathsf{S}_{d-1}} \frac{|L_t(x,\theta) \cap A|}{|L_t(x,\theta)|} \, \mathrm{d}\sigma(\theta) \, \mathrm{d}t,$$

where $|\cdot|$ here denotes the one-dimensional Lebesgue measure. This modification is tempting since the uniform hit-and-run algorithm is, at least in some scenarios, implementable. For example, when the super-level sets are convex (ρ is quasi-conave), using a bisection method we can roughly approximate the intersection points of the line $x + r\theta$ with $K(t)$ and use a one-dimensional acceptance/rejection approach to sample the uniform distribution on $L_t(x,\theta)$. Further, the simple slice sampler and the hybrid slice sampler are reversible with respect to π; see, for example, [15].

2.2.3. *Random walk Metropolis.* The random walk Metropolis in $\mathbb{R}^d$ provides an easy-to-implement method for Markov chain sampling. Further, it is well studied and different convergence results are known; see, for example, [11], [21], and [33].

To guarantee that a certain operator is positive, we consider a *lazy version of a random walk Metropolis*. Further, we assume in the following that $q : \mathbb{R}^d \to [0, \infty)$ is a rotational invariant probability density on $\mathbb{R}^d$, i.e. $q(r\theta_1) = q(r\theta_2)$ for $\theta_1, \theta_2 \in \mathsf{S}_{d-1}$. The rotational invariance guarantees that q is symmetric. We now provide two examples which satisfy the rotational invariance.

Example 1. Let $x \in \mathbb{R}^d$, $\delta > 0$, and B_d be the Euclidean unit ball with $\kappa_d := \text{vol}_d(\mathsf{B}_d)$. Then we can set $q(x) := \mathbf{1}_{\mathsf{B}_d}(\delta^{-1}x)/(\delta^d \kappa_d)$ and the corresponding random walk Metropolis is known as the δ-ball walk.

Example 2. Again, let $x \in \mathbb{R}^d$ and set $q(x) := \exp(-\frac{1}{2}|x|^2)/(2\pi)^{d/2}$. The corresponding random walk Metropolis is known as the Gaussian random walk.

In Example 1, the proposal q depends on a parameter δ. In connection to this, we note that in recent years optimal scaling results concerning a parameter of the proposal of the random walk Metropolis algorithm has attracted a lot of attention; see [27], [31] and [34].

In Algorithm 4, we describe a single transition of the (lazy) random walk Metropolis algorithm with proposal q.

Algorithm 4. (*Random walk Metropolis.*) Transition from current state x to next state y with proposal q.

1. Sample $z \sim q$ and $u_1, u_2 \sim \text{uniform}(0, 1)$ independently.
2. If $x + z \in K$, $u_1 \le \frac{1}{2}$, and $u_2 < \min\{1, \rho(x+z)/\rho(x)\}$ then accept the proposal and set $y := x + z$, otherwise reject the proposal and set $y := x$.

To simplify the notation, we define the acceptance probability of a proposed state $y = x + z$ as

$$\alpha(x, y) = \begin{cases} \min\left\{1, \dfrac{\rho(y)}{\rho(x)}\right\} & \text{for } x, y \in K, \\ 0 & \text{otherwise.} \end{cases}$$

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

Then, the transition kernel is

$$M(x, A) = \frac{1}{2} \int_K \mathbf{1}_A(y)\alpha(x, y)q(y - x)\,\mathrm{d}y \quad \text{for } A \subseteq K$$

with $x \notin A$ and $M(x, \{x\}) = 1 - M(x, K \setminus \{x\})$.

3. Auxiliary variable Markov chains

We develop a systematic approach to compare Markov chains which can be described by a suitable two-step procedure. The fact that many Markov chains are of this form goes back to [1] and the idea of a comparison of this type was developed in [40] and [41] in a specific setting.

For this, let $\mathcal{A}$ be an arbitrary (index) set equipped with a σ-finite measure λ. By assumption, we have a function $s\colon K \times \mathcal{A} \to [0, \infty)$ which satisfies:

- for all $x \in K$, $s(x, \cdot)$ is a probability density function according to λ;
- for all $a \in \mathcal{A}$, $s(\cdot, a)$ is integrable according to π.

Define, for almost all $a \in \mathcal{A}$ (concerning λ), a probability measure π_a on K induced by $s(x, a)$, i.e.

$$\pi_a(A) = \frac{\int_A s(x, a)\,\mathrm{d}\pi(x)}{\int_K s(x, a)\,\mathrm{d}\pi(x)}, \qquad A \subseteq K.$$

In addition, we assume that

- for every $a \in \mathcal{A}$, we have an equivalence relation $\sim_a$ and by $[x]_a = \{y \in K : x \sim_a y\}$, we denote the equivalence class of K with respect to $\sim_a$ to which x belongs;
- for (λ-almost) every $a \in \mathcal{A}$, we have a transition kernel P_a on $(K, \mathcal{B}(K))$ such that $P_a(x, A) = 0$ for each $x \in K$ and $A \subseteq K \setminus [x]_a$.

With this we can define a Markov chain in K for which a single transition, starting from $x \in K$, can be written as the following procedure.

1. Sample $a \in \mathcal{A}$ according to the distribution with density $s(x, \cdot)$.
2. Generate the next state with respect to $P_a(x, \cdot)$.

That is, the transition kernel P is

$$P(x, A) = \int_{\mathcal{A}} P_a(x, A)s(x, a)\,\mathrm{d}\lambda(a) \quad \text{for } A \subseteq K. \tag{1}$$

Remark 1. Since the support of the measure $P_a(x, \cdot)$ is contained in $[x]_a$ and $[x]_a = [y]_a$ if $y \in [x]_a$, we can interpret a Markov chain with transition kernel P_a on K and initial state x as a Markov chain on $[x]_a$.

Clearly, for every Markov chain a transition can be written in this form. (Simply, set $P_a(x, \cdot) := P(x, \cdot)$ and take $[x]_a \equiv K$ for arbitrary $\mathcal{A}$ and λ.) However, there might also be more interesting equivalence relations and scenarios as we now illustrate for the hit-and-run transition kernel.

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

Example 3. Here, let $\mathcal{A} := \mathsf{S}_{d-1}$ and $\lambda := \sigma$. Then, for every $a \in \mathcal{A}$, let $x \sim_a y$ if and only if $x \in L(y, a)$ such that $[x]_a := L(x, a)$. Further, for $(x, a) \in K \times \mathcal{A}$, set $s(x, a) := 1$, which implies $\pi_a = \pi$ for all $a \in \mathcal{A}$. For (λ-almost every) $a \in \mathcal{A}$, it follows that

$$H_a(x, A) = \frac{\int_{[x]_a} \rho(y) \mathbf{1}_A(y) \, \mathrm{d}y}{\int_{[x]_a} \rho(y) \, \mathrm{d}y}, \qquad A \subseteq K,$$

is a transition kernel on K and we can write the Markov operator $H \colon L_2(\pi) \to L_2(\pi)$ of the hit-and-run algorithm in the form of (1) as

$$Hf(x) = \int_{\mathcal{A}} \int_K f(y) H_a(x, \, \mathrm{d}y) \, \mathrm{d}\lambda(a).$$

In some cases it is even possible to represent the transitions of two different Markov chains in the form of (1) with the same equivalence relations $\sim_a$ and $s(x, a)$, so that the corresponding 'inner' transition kernels are much easier to analyze. We show that a suitable relation of these inner kernels is enough to compare the original Markov chains.

Lemma 1. *Assume that for (λ-almost) all $a \in \mathcal{A}$, there are transition kernels $P_a^{(1)}$, $P_a^{(2)}$ on K such that*

(i) *$P_a^{(1)}$ and $P_a^{(2)}$ are self-adjoint operators on $L_2(\pi_a)$;*

(ii) *$P_a^{(1)}$ is positive on $L_2(\pi_a)$;*

(iii) *$P_a^{(1)} P_a^{(2)} = P_a^{(2)}$.*

Then, for the operators $P_1, P_2 \colon L_2(\pi) \to L_2(\pi)$ defined by

$$P_i f(x) = \int_{\mathcal{A}} P_a^{(i)} f(x) s(x, a) \, \mathrm{d}\lambda(a), \qquad i \in \{1, 2\},$$

it holds that $P_1 \leq P_2$.

Proof. First, we show that P_i can be written as a product of suitable operators. In order to do so, let μ be a probability measure on $K \times \mathcal{A}$ given by

$$\mu(B) := \int_K \int_{\mathcal{A}} \mathbf{1}_B(x, a) s(x, a) \, \mathrm{d}\lambda(a) \, \mathrm{d}\pi(x) \quad \text{for } B \subseteq K \times \mathcal{A},$$

and let $L_2(\mu)$ be the Hilbert space of functions $g \colon K \times \mathcal{A} \to \mathbb{R}$ with finite norm $\|g\|_\mu = \langle g, g \rangle_\mu^{1/2}$, where the inner product of $g, h \in L_2(\mu)$ is defined by

$$\langle g, h \rangle_\mu = \int_{K \times \mathcal{A}} g(x, a) h(x, a) \, \mathrm{d}\mu(x, a) = \int_K \int_{\mathcal{A}} g(x, a) h(x, a) s(x, a) \, \mathrm{d}\lambda(a) \, \mathrm{d}\pi(x).$$

Further, let $R \colon L_2(\mu) \to L_2(\pi)$ be the operator given by

$$Rg(x) = \int_{\mathcal{A}} g(x, a) s(x, a) \, \mathrm{d}\lambda(a), \qquad g \in L_2(\mu).$$

Since

$$\langle f, Rg \rangle_\pi = \int_K \int_{\mathcal{A}} f(x) g(x, a) s(x, a) \, \mathrm{d}\lambda(a) \, \mathrm{d}\pi(x),$$

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

we see that the adjoint operator $R^*\colon L_2(\pi) \to L_2(\mu)$ satisfies

$$R^* f(x,a) = f(x), \qquad f \in L_2(\pi) \quad \text{for } a \in \mathcal{A},\, x \in K.$$

It is useful to write $g_a(x) = g(x,a)$ for $g \in L_2(\mu)$ and fixed $a \in \mathcal{A}$. Clearly, $g_a \in L_2(\pi_a)$ for almost every $a \in \mathcal{A}$ (with respect to λ) such that $P_a^{(i)} g_a$ is well defined. Let

$$\widetilde{P}_i g(x,a) = \int_K g(y,a) P_a^{(i)}(x,\, \mathrm{d}y), \qquad g \in L_2(\mu),$$

and note that the operator satisfies $\widetilde{P}_i\colon L_2(\mu) \to L_2(\mu)$. An immediate consequence of the definitions is

$$P_i = R \widetilde{P}_i R^*, \qquad i \in \{1,2\}.$$

We prove different properties of $\widetilde{P}_1$ and $\widetilde{P}_2$. For $i \in \{1,2\}$, we show that $\widetilde{P}_i$ is self-adjoint on $L_2(\mu)$. Note that, by our assumptions, we know that $P_a^{(i)}$ is self-adjoint on $L_2(\pi_a)$. Setting

$$C_a = \int_K s(x,a)\, \mathrm{d}\pi(x),$$

we have, for $g, h \in L_2(\mu)$,

$$\begin{aligned}
\langle \widetilde{P}_i g, h\rangle_\mu &= \int_K \int_{\mathcal{A}} \int_K g(y,a) P_a^{(i)}(x,\, \mathrm{d}y) h(x,a) s(x,a)\, \mathrm{d}\lambda(a)\, \mathrm{d}\pi(x) \\
&= \int_{\mathcal{A}} C_a \int_K \int_K g(y,a) P_a^{(i)}(x,\, \mathrm{d}y) h(x,a)\, \mathrm{d}\pi_a(x)\, \mathrm{d}\lambda(a) \\
&= \int_{\mathcal{A}} C_a \langle P_a^{(i)} g_a, h_a\rangle_{\pi_a}\, \mathrm{d}\lambda(a) \\
&= \int_{\mathcal{A}} C_a \langle g_a, P_a^{(i)} h_a\rangle_{\pi_a}\, \mathrm{d}\lambda(a) \\
&= \langle g, \widetilde{P}_i h\rangle_\mu.
\end{aligned}$$

By the same line of reasoning, for $g \in L_2(\mu)$, we have

$$\langle \widetilde{P}_1 g, g\rangle_\mu = \int_{\mathcal{A}} C_a \langle P_a^{(1)} g_a, g_a\rangle_{\pi_a}\, \mathrm{d}\lambda(a) \geq 0$$

such that $\widetilde{P}_1$ preserves the positivity of $P_a^{(1)}$, i.e. $\widetilde{P}_1$ is positive on $L_2(\mu)$.

Further, for $g \in L_2(\mu)$, we obtain

$$\widetilde{P}_1 \widetilde{P}_2 g(x,a) = P_a^{(1)} P_a^{(2)} g_a(x) = P_a^{(2)} g_a(x) = \widetilde{P}_2 g(x,a).$$

By the self-adjointness of $\widetilde{P}_1$ and $\widetilde{P}_2$, we also have $\widetilde{P}_2 \widetilde{P}_1 = \widetilde{P}_2$.

Now we gather all our results to prove the assertion. By positivity, we know that $\widetilde{P}_1$ has a unique positive square root N, i.e. $N^2 = \widetilde{P}_1$. It is well known that N commutes with every operator that commutes with $\widetilde{P}_1$ (see, for example, [14, Theorem 9.4-2]), in particular, $N\widetilde{P}_2 = \widetilde{P}_2 N$. We obtain

$$\begin{aligned}
\langle P_2 f, f\rangle_\pi &= \langle R \widetilde{P}_2 R^* f, f\rangle_\pi \\
&= \langle \widetilde{P}_2 R^* f, R^* f\rangle_\mu \\
&= \langle \widetilde{P}_1 \widetilde{P}_2 R^* f, R^* f\rangle_\mu
\end{aligned}$$

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

$$
\begin{aligned}
&= \langle \widetilde{P}_2 N R^* f, N R^* f \rangle_\mu \\
&\le \langle N R^* f, N R^* f \rangle_\mu \\
&= \langle \widetilde{P}_1 R^* f, R^* f \rangle_\mu \\
&= \langle P_1 f, f \rangle_\pi ,
\end{aligned}
$$

where the inequality is due to $\|\widetilde{P}_2\| \le 1$, which holds since $\widetilde{P}_2$ is a Markov operator. □

Thus, for each of the intended comparisons, say between P_1 and P_2, we have to

- find representations

$$P_i f(x) = \int_{\mathcal{A}} P_a^{(i)} f(x) s(x, a) \, \mathrm{d}\lambda(a), \qquad i \in \{1, 2\};$$

- check assumptions (i)–(iii) of Lemma 1, i.e. reversibility of $P_a^{(i)}$ with respect to π_a, positivity of $P_a^{(1)}$, and $P_a^{(1)} P_a^{(2)} = P_a^{(2)}$.

Remark 2. Let us comment on the necessity and possible generalizations of the assumptions in Lemma 1. The second assumption, i.e. the positivity of $P_a^{(1)}$, is most likely an artifact of the proof technique and we conjecture that the result holds without this assumption. However, in the proof it is essential and we are not able to remove it. The same problem already appeared in [40] and [41]. The third assumption of the lemma says, roughly speaking, that a step of the Markov chain with corresponding operator $P_a^{(1)}$ 'cannot be seen' if followed by $P_a^{(2)}$. As we see in the final steps of the proof, this could be replaced, for example, by the weaker assumption that $\widetilde{P}_1 \le \widetilde{P}_2$, i.e. $\langle \widetilde{P}_2 g, g \rangle_\mu \le \langle \widetilde{P}_1 g, g \rangle_\mu$ for all $g \in L_2(\mu)$. However, at least in the examples we consider, the relatively easy-to-check assumption of Lemma 1(iii) is already fulfilled.

4. Main result

In this section we apply Lemma 1 to prove the following theorem.

Theorem 1. *Let M, U, H, S be the Markov operators of the (lazy) random walk Metropolis algorithm with rotational invariant proposal q, the hybrid slice sampler based on hit-and-run, the hit-and-run algorithm, and the simple slice sampler. Then*

$$M \le U \le H, \qquad M \le U \le S.$$

Before proving the different inequalities in the statement, we start by recalling some notions and state a useful lemma. Recall that S_{d-1} denotes the Euclidean unit sphere in $\mathbb{R}^d$ and σ is the uniform distribution on S_{d-1}. For $\theta \in \mathsf{S}_{d-1}$, $x \in \mathbb{R}^d$, and $t \in [0, \infty)$, let

$$K(t) = \{x \in K \mid \rho(x) > t\}$$

be the super-level set of ρ determined by t, and

$$L_t(x, \theta) = \{x + \theta r \in K(t) \mid r \in \mathbb{R}\}$$

be the chord in $K(t)$ through x in direction θ. Moreover, define $L(x, \theta) \coloneqq L_0(x, \theta)$ and, for a set $A \subseteq K = K(0)$, let

$$\Pi_\theta(A) = \{y \in \mathbb{R}^d \mid y \perp \theta, L(y, \theta) \cap A \ne \varnothing\}$$

be the orthogonal projection of A to the hyperplane that is orthogonal to θ. We obtain the following useful results by an application of Fubini's theorem and the integral transformation to polar coordinates.

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

Lemma 2. *Let $t \geq 0$ and $\theta \in \mathsf{S}_{d-1}$. For Lebesgue integrable $f\colon K(t) \to \mathbb{R}$, we have*

$$\int_{K(t)} f(x)\,\mathrm{d}x = \int_{\Pi_\theta(K(t))} \int_{L_t(x,\theta)} f(y)\,\mathrm{d}y\,\mathrm{d}x, \tag{2}$$

and for fixed $x \in \mathbb{R}^d$, it holds that

$$\int_{K(t)} f(y)\,\mathrm{d}y = \frac{d\kappa_d}{2} \int_{\mathsf{S}_{d-1}} \int_{L_t(x,\theta)} f(y)|x-y|^{d-1}\,\mathrm{d}y\,\mathrm{d}\sigma(\theta), \tag{3}$$

where $\kappa_d = \mathrm{vol}_d(\mathsf{B}_d)$ denotes the volume of the d-dimensional Euclidean unit ball.

Note that in both identities the inner integral on the right-hand side is over a one-dimensional subset of $\mathbb{R}^d$ and the integration is with respect to the one-dimensional Lebesgue measure.

The different inequalities in Theorem 1 will be proved in the following sections where we use the notation of Section 3.

4.1. Hit-and-run versus hybrid slice sampler

For the hit-and-run we use the scenario described in Example 3. Hence, let $\mathcal{A} := \mathsf{S}_{d-1}$ and $\lambda := \sigma$. Further, for $(x, a) \in K \times \mathcal{A}$, set $s(x, a) := 1$ such that $[x]_a := L_0(x, a)$. This implies $\pi_a = \pi$ for all $a \in \mathcal{A}$. From

$$H_a(x, A) = \frac{\int_{[x]_a} \rho(y)\mathbf{1}_A(y)\,\mathrm{d}y}{\int_{[x]_a} \rho(y)\,\mathrm{d}y}, \qquad A \subseteq K,$$

we can write the Markov operator $H\colon L_2(\pi) \to L_2(\pi)$ of the hit-and-run algorithm as

$$Hf(x) = \int_{\mathcal{A}} \int_K f(y) H_a(x, \mathrm{d}y)\,\mathrm{d}\lambda(a).$$

Let

$$U_a(x, A) = \frac{1}{\rho(x)} \int_0^{\rho(x)} \int_{L_t(x,a)} \frac{\mathbf{1}_A(y)}{|L_t(x, a)|}\,\mathrm{d}y\,\mathrm{d}t$$

and observe that $L_t(x, a) = [x]_a \cap K(t)$. With this we can write the Markov operator $U\colon L_2(\pi) \to L_2(\pi)$ of the hybrid slice sampler based on hit-and-run as

$$Uf(x) = \int_A \int_K f(y) U_a(x, \mathrm{d}y)\,\mathrm{d}\lambda(a).$$

Thus, we have a common representation. Now we can check assumptions (i)–(iii) of Lemma 1.

(i) *The reversibility of H_a and U_a with respect to π_a.* We show only the reversibility of U_a, since the reversibility of H_a follows by the same line of reasoning. For $A, B \subseteq K$, it is enough to prove that

$$\int_A U_a(x, B)\rho(x)\,\mathrm{d}x = \int_B U_a(x, A)\rho(x)\,\mathrm{d}x. \tag{4}$$

By an application of (2) with $t = 0$, the equality $\mathbf{1}_{[0,\rho(y))}(t) = \mathbf{1}_{K(t)}(y)$, and the fact that $y \in [x]_a$ implies $L_t(y, a) = L_t(x, a)$, we obtain

$$\begin{aligned}\int_A U_a(x, B)\rho(x)\,\mathrm{d}x &= \int_{\Pi_a(K)} \int_{[x]_a} \mathbf{1}_A(y) U_a(y, B)\rho(y)\,\mathrm{d}y\,\mathrm{d}x\end{aligned}$$

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

$$= \int_{\Pi_a(K)} \int_{[x]_a} \mathbf{1}_A(y) \int_0^{\rho(y)} \int_{L_t(y,a)} \frac{\mathbf{1}_B(z)}{|L_t(y,a)|} \,\mathrm{d}z \,\mathrm{d}t \,\mathrm{d}y \,\mathrm{d}x$$
$$= \int_{\Pi_a(K)} \int_0^{\infty} \frac{1}{|L_t(x,a)|} |L_t(x,a) \cap A| |L_t(x,a) \cap B| \,\mathrm{d}t \,\mathrm{d}x.$$

Observe that on the right-hand side, A and B are interchangeable which proves (4).

(ii) *The positivity of U_a.* By similar arguments as in the proof of (4), we obtain, with $c = \int_K \rho(x) \,\mathrm{d}x$,

$$\langle U_a f, f \rangle_\pi = \frac{1}{c} \int_{\Pi_a(K)} \int_0^{\infty} \frac{1}{|L_t(x,a)|} \left(\int_{L_t(x,a)} f(y) \,\mathrm{d}y \right)^2 \mathrm{d}t \,\mathrm{d}x \geq 0.$$

(iii) *On $U_a H_a = H_a$.* From the definition of the Markov kernels H_a, it is obvious that $y \in [x]_a$ implies $H_a(y, A) = H_a(x, A)$ for all $A \subseteq K$. This implies that

$$U_a H_a(x, A) = \frac{1}{\rho(x)} \int_0^{\rho(x)} \int_{L_t(x,a)} H_a(y, A) \frac{\mathrm{d}y}{|L_t(x,a)|} \,\mathrm{d}t = H_a(x, A)$$

and, hence, $U_a H_a = H_a$.

Thus, as a direct consequence of Lemma 1, we obtain $U \leq H$.

4.2. Simple versus hybrid slice sampler

Here, we derive another representation of the hybrid slice sampler adapted to simple slice sampling.

To do so, let $\mathcal{A} := [0, \infty)$ and λ be the one-dimensional Lebesgue measure. Further, for $(x, a) \in K \times \mathcal{A}$, set $s(x, a) := \mathbf{1}_{[0,\rho(x))}(a)/\rho(x)$ such that $[x]_a := K(a)$. Observe that

$$\pi_a(A) = \frac{\mathrm{vol}_d(A \cap K(a))}{\mathrm{vol}_d(K(a))}.$$

For any $x \in K$, let

$$S_a(x, A) = \pi_a(A), \qquad A \subseteq K.$$

Clearly, the Markov operator $S \colon L_2(\pi) \to L_2(\pi)$ of the simple slice sampler can be written as

$$Sf(x) = \int_{\mathcal{A}} \int_K f(y) S_a(x, \mathrm{d}y) s(x, a) \,\mathrm{d}\lambda(a).$$

Note that, with this notation, we have $L_a(x, \theta) = [x]_a \cap L(x, \theta)$ and

$$U_a(x, A) = \int_{\mathbb{S}_{d-1}} \int_{L_a(x,\theta)} \frac{\mathbf{1}_A(y)}{|L_a(x,\theta)|} \,\mathrm{d}y \,\mathrm{d}\sigma(\theta).$$

With this we can write the Markov operator $U \colon L_2(\pi) \to L_2(\pi)$ of the hybrid slice sampler based on hit-and-run as

$$Uf(x) = \int_{\mathcal{A}} \int_K f(y) U_a(x, \mathrm{d}y) s(x, a) \,\mathrm{d}\lambda(a).$$

Again, we have a common representation and it remains to check assumptions (i)–(iii) of Lemma 1.

(i) *The reversibility of S_a and U_a with respect to π_a.* Since $S_a(x, A) = \pi_a(A)$ the reversibility of S_a is obvious. Observe that π_a is the uniform distribution on $[x]_a = K(a)$ and

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

$U_a(x,\cdot)$ performs a uniform hit-and-run step on $K(a)$, which is known to be reversible; see, for example, [35, Lemma 4.10].

(ii) *The positivity of U_a.* With $c = \mathrm{vol}_d(K(a))$, by an application of (2) and by the fact that $y \in L_a(x,\theta)$ implies $L_a(x,\theta) = L_a(y,\theta)$, it holds that

$$\begin{aligned}\langle U_a f, f\rangle_{\pi_a} &= \frac{1}{c}\int_{K(a)}\int_{\mathsf{S}_{d-1}}\int_{L_a(x,\theta)}\frac{f(z)f(x)}{|L_a(x,\theta)|}\,\mathrm{d}z\,\mathrm{d}\sigma(\theta)\,\mathrm{d}x\\ &= \frac{1}{c}\int_{\mathsf{S}_{d-1}}\int_{\Pi_\theta(K(a))}\int_{L_a(x,\theta)}\int_{L_a(y,\theta)}\frac{f(y)f(z)}{|L_a(y,\theta)|}\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}\sigma(\theta)\\ &= \frac{1}{c}\int_{\mathsf{S}_{d-1}}\int_{\Pi_\theta(K(a))}\frac{1}{|L_a(x,\theta)|}\left(\int_{L_a(x,\theta)} f(y)\,\mathrm{d}y\right)^2\mathrm{d}x\,\mathrm{d}\sigma(\theta)\\ &\geq 0.\end{aligned}$$

(iii) *On $U_a S_a = S_a$.* This follows immediately from $S_a = \pi_a$ and the reversibility of U_a with respect to π_a.

Thus, as a direct consequence of Lemma 1, we obtain $U \leq S$.

4.3. Hybrid slice sampler versus Metropolis

Again we need a suitable representation for the hybrid slice sampler based on hit-and-run.

Here, let $\mathcal{A} := \mathsf{S}_{d-1}\times[0,\infty)$ and λ be the product measure of σ and the one-dimensional Lebesgue measure. Further, for $x \in K$ and $(a_1,a_2) \in \mathcal{A}$, set $s(x,a_1,a_2) = \mathbf{1}_{[0,\rho(x)]}(a_2)/\rho(x)$ such that $[x]_{(a_1,a_2)} := L_{a_2}(x,a_1)$. Observe that

$$\pi_{(a_1,a_2)}(A) = \frac{\mathrm{vol}_d(A\cap K(a_2))}{\mathrm{vol}_d(K(a_2))}, \qquad A \subseteq K,$$

is the uniform distribution in $K(a_2)$. From

$$U_{(a_1,a_2)}(x,A) = \int_{[x]_{(a_1,a_2)}} \mathbf{1}_A(y)\frac{\mathrm{d}y}{|[x]_{(a_1,a_2)}|},$$

we have a representation of the Markov operator $U\colon L_2(\pi)\to L_2(\pi)$ of the hybrid slice sampler, i.e.

$$Uf(x) = \int_{\mathcal{A}}\int_K f(y)U_a(x,\mathrm{d}y)s(x,a)\,\mathrm{d}\lambda(a)$$

with $a \in \mathcal{A}$. Now we have to represent the random walk Metropolis in the same fashion. To do so, we define

$$\eta(x,y) = \tfrac{1}{2}d\kappa_d q(y-x)|y-x|^{d-1}, \qquad x,y\in\mathbb{R}^d,$$

and, for $A \subseteq K$, let

$$M_a(x,A) = \frac{1}{2}\int_{[x]_a}\mathbf{1}_A(y)\eta(x,y)\,\mathrm{d}y + \mathbf{1}_A(x)\left(1-\frac{1}{2}\int_{[x]_a}\eta(x,y)\,\mathrm{d}y\right).$$

Here it is essential that q is rotational invariant; namely, this property ensures that

$$\int_{[x]_a}\eta(x,y)\,\mathrm{d}y \leq 1.$$

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

Note that M_a is again a lazy transition kernel, since $M_a(x, \{x\}) \geq \frac{1}{2}$. For $x \notin A$, from (3) we have

$$\begin{aligned} M(x, A) &= \frac{1}{2} \int_K \mathbf{1}_A(y)\alpha(x, y)q(y - x)\,\mathrm{d}y \\ &= \frac{1}{2\rho(x)} \int_0^{\rho(x)} \int_K \mathbf{1}_{[0,\rho(y)]}(t)\mathbf{1}_A(y)q(y - x)\,\mathrm{d}y\,\mathrm{d}t \\ &= \frac{1}{2\rho(x)} \int_0^{\rho(x)} \int_{K(t)} \mathbf{1}_A(y)q(y - x)\,\mathrm{d}y\,\mathrm{d}t \\ &= \frac{d\kappa_d}{4\rho(x)} \int_{\mathsf{S}_{d-1}} \int_0^{\rho(x)} \int_{L_t(x,\theta)} \mathbf{1}_A(y)q(x - y)|x - y|^{d-1}\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}\sigma(\theta) \\ &= \int_{\mathcal{A}} M_a(x, A)s(x, a)\,\mathrm{d}\lambda(a). \end{aligned}$$

Thus, the transition kernel of the random walk Metropolis has the desired representation. It remains to check conditions (i)–(iii) of Lemma 1.

(i) *The reversibility of M_a and U_a with respect to π_a.* This follows again by a suitable application of (2). Let $a = (a_1, a_2)$. Due to its simple form, the reversibility of U_a is obvious. It is enough to show that, for disjoint $A, B \subseteq K$,

$$\int_{A\cap K(a_2)} M_a(x, B)\,\mathrm{d}x = \int_{B\cap K(a_2)} M_a(x, A)\,\mathrm{d}x.$$

Since $L_{a_2}(x, a_1) = L_{a_2}(y, a_1)$, if $y \in L_{a_2}(x, a_1)$ we have

$$\begin{aligned} &\frac{1}{2} \int_{A\cap K(a_2)} \int_{[x]_a} \mathbf{1}_B(y)\eta(x, y)\,\mathrm{d}y\,\mathrm{d}x \\ &\quad= \frac{1}{2} \int_{K(a_2)} \int_{L_{a_2}(x,a_1)} \mathbf{1}_A(x)\mathbf{1}_B(y)\eta(x, y)\,\mathrm{d}y\,\mathrm{d}x \\ &\quad= \frac{1}{2} \int_{\Pi_{a_1}(K(a_2))} \int_{L_{a_2}(x,a_1)} \int_{L_{a_2}(y,a_1)} \mathbf{1}_A(y)\mathbf{1}_B(z)\eta(y, z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x \\ &\quad= \frac{1}{2} \int_{\Pi_{a_1}(K(a_2))} \int_{L_{a_2}(x,a_1)} \int_{L_{a_2}(x,a_1)} \mathbf{1}_A(y)\mathbf{1}_B(z)\eta(y, z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x. \end{aligned}$$

By the symmetry of$q$, i.e. $q(y - z) = q(z - y)$, we obtain $\eta(y, z) = \eta(z, y)$. This leads to the reversibility.

(ii) *The positivity of M_a.* From the fact that M_a is a lazy transition kernel we have positivity.

(iii) *On $M_aU_a = U_a$.* From the fact that $y \in [x]_a$ implies $[x]_a = [y]_a$ and, hence, $U_a(x, A) = U_a(y, A)$ for all $A \subseteq K$, we obtain

$$\begin{aligned} M_aU_a(x, A) &= \frac{1}{2} \int_{[x]_a} U_a(y, A)\eta(x, y)\,\mathrm{d}y + U_a(x, A)\left(1 - \frac{1}{2} \int_{[x]_a} \eta(x, y)\,\mathrm{d}y\right) \\ &= U_a(x, A). \end{aligned}$$

Thus, as a direct consequence of Lemma 1, we obtain $M \leq U$.

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

5. Concluding remarks

In this paper we have presented a technique to compare the efficiency of Markov chains of a specific type. Using this technique, we provided two comparison hierarchies according to a partial ordering of Markov operators of four prominent Markov chains for sampling general distributions in $\mathbb{R}^d$. The comparison with respect to the partial ordering led to comparison results according to different criteria, for example the spectral gap, the conductance, or the log-Sobolev constant; see Section 2.1.1.

Let us mention here that the computational cost for the simulation of each individual Markov chain was not taken into account. There seems to be a trade-off between efficiency and computational cost which needs further investigation. We leave this open for future work.

Finally, there are three open problems related to the considered Markov chains.

First is the relationship between the hit-and-run algorithm and the simple slice sampler. It is easy to see that there cannot be a general result as in the other cases. For this, consider two examples.

- If π is the uniform distribution on K then one step of the simple slice sampler is enough to sample π, while for the hit-and-run algorithm it is not (as long as $d > 1$). Hence, in this situation, hit-and-run is worse.
- Let $d = 1$. Then the hit-and-run algorithm samples π in one step, regardless of π. Hence, in this situation, hit-and-run is better.

It is of interest to find cases where hit-and-run is better. This is because, in general, the computational cost of the simple slice sampler is (if it is at all implementable) much higher.

The second open problem concerns reverse inequalities. That is, if a Markov chain is better than another, how much better is it? This seems to be a delicate question and we can answer it only for toy examples. The techniques that were used in [41] in a discrete setting do not seem to work in our setting.

The final problem we want to mention is whether there exists a similar hierarchy for the *mixing time*. That is, the number of steps that are needed to make the total variation distance 'small'; see Section 2.1.1. Certainly, the answer to this question depends additionally on the choice of the initial distribution. But the (quite analytical) techniques of this paper (see also [2], [15], [29], and [41] do not seem to be suitable for this purpose. One interesting approach in this direction for Markov chains on discrete state spaces can be found in [9].

References

[1] Andersen, C. and Diaconis, P. (2007). Hit and run as a unifying device. *J. Soc. Fr. Stat.* **148,** 5–28.

[2] Andrieu, C. and Vihola, M. (2016). Establishing some order amongst exact approximations of MCMCs. *Ann. Appl. Prob.* **26,** 2661–2696.

[3] Baxendale, P. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Prob.* **15,** 700–738.

[4] Bélisle, C., Romeijn, H. and Smith, R. (1993). Hit-and-run algorithms for generating multivariate distributions. *Math. Operat. Res.* **18,** 255–266.

[5] Brooks, S., Gelman, A., Jones, G. and Meng, X. (eds) (2011). *Handbook of Markov chain Monte Carlo*. Chapman and Hall, Boca Raton.

[6] Casella, G. and Robert, C. (2004). *Monte Carlo Statistical Methods*, 2nd edn. Springer, New York.

[7] Chen, M. (2005). Eigenvalues, inequalities, and ergodic theory. In *Probability and its Applications*, Springer, London.

[8] Diaconis, P. and Freedman, D. (1997). On the hit and run process. Tech. Rep. University of California **497**, 1–16.

https://doi.org/10.1017/jpr.2018.78 Published online by Cambridge University Press

[9] Fill, J. and Kahn, J. (2013). Comparison inequalities and fastest mixing Markov chains. *Ann. App. Prob.* **23**, 1778–1816.
[10] Higdon, D. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *J. Amer. Statist. Assoc.* **93,** 585–595.
[11] Jarner, S. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stoch. Process. Appl.* **85,** 341–361.
[12] Kaufman, D. and Smith, R. (1998). Direction choice for accelerated convergence in hit-and-run sampling. *Operat. Res.* **46,** 84–95.
[13] Kiatsupaibul, S., Smith, R. and Zabinsky, Z. (2011). An analysis of a variation of hit-and-run for uniform sampling from general regions. *ACM Trans. Model. Comput. Simul.* **21,** 1–11.
[14] Kreyszig, E. (1989). *Introductory Functional Analysis with Applications*. John Wiley, New York.
[15] Latuszyński, K. and Rudolf, D. (2014). Convergence of hybrid slice sampling via spectral gap. Preprint. Available at https://arxiv.org/abs/1409.2709.
[16] Lovász, L. (1999). Hit-and-run mixes fast. *Math. Program.* **86,** 443–461.
[17] Lovász, L. and Vempala, S. (2006). Fast algorithms for logconcave functions: sampling, rounding, integration and optimization. In *Proc. 47th Annual IEEE Symposium on Foundations of Computer Science*, Berkeley, CA, pp. 57–68.
[18] Lovász, L. and Vempala, S. (2006). Hit-and-run from a corner. *SIAM J. Comput.* **35,** 985–1005.
[19] Lovász, L. and Vempala, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms* **30,** 307–358.
[20] Mathé, P. and Novak, E. (2007). Simple Monte Carlo and the Metropolis algorithm. *J. Complexity* **23,** 673–696.
[21] Mengersen, K. and Tweedie, R. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24,** 101–121.
[22] Mira, A. (2001). Ordering and improving the performance of Monte Carlo Markov chains. *Statist. Sci.* **16,** 340–350.
[23] Mira, A. and Leisen, F. (2009). Covariance ordering for discrete and continuous time Markov chains. *Statistica Sinica* **19,** 651–666.
[24] Mira, A. and Tierney, L. (2002). Efficiency and convergence properties of slice samplers. *Scand. J. Statist.* **29,** 1–12.
[25] Mira, A., Møller, J. and Roberts, G. (2001). Perfect slice samplers. *J. R. Statist. Soc. B* **63,** 593–606.
[26] Montenegro, R. and Tetali, P. (2006). Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.* **1,** 121 pp.
[27] Neal, P., Roberts, G. and Yuen, W. (2012). Optimal scaling of random walk Metropolis algorithms with discontinuous target densities. *Ann. Appl. Prob.* **22,** 1880–1927.
[28] Neal, R. (2003). Slice sampling. *Ann. Statist.* **31,** 705–767.
[29] Peskun, P. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60,** 607–612.
[30] Roberts, G. and Rosenthal, J. (1999). Convergence of slice sampler Markov chains. *J. R. Statist. Soc. B* **61,** 643–660.
[31] Roberts, G. and Rosenthal, J. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.* **16,** 351–367.
[32] Roberts, G. and Rosenthal, J. (2002). The polar slice sampler. *Stoch. Models* **18,** 257–280.
[33] Roberts, G. and Tweedie, R. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83,** 95–110.
[34] Roberts, G., Gelman, A. and Gilks, W. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Prob.* **7,** 110–120.
[35] Rudolf, D. (2012). Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Math.* **485,** 93 pp.
[36] Rudolf, D. (2013). Hit-and-run for numerical integration. In *Monte Carlo and Quasi-Monte Carlo Methods 2012* (Springer Proc. Math. Statist. **65**), eds J. Dick, F. Y. Kuo, G. Peters, and I. H. Sloan. Springer, Berlin, pp. 597–612.
[37] Rudolf, D. and Ullrich, M. (2013). Positivity of hit-and-run and related algorithms. *Electron. Commun. Prob.* **18,** 1–8.
[38] Smith, R. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operat. Res.* **32,** 1296–1308.
[39] Tierney, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Prob.* **8,** 1–9.
[40] Ullrich, M. (2012). Rapid mixing of Swendsen-Wang and single-bond dynamics in two dimensions. Preprint. Available at https://arxiv.org/abs/1202.6321.
[41] Ullrich, M. (2014). Rapid mixing of Swendsen–Wang dynamics in two dimensions. *Dissertationes Math.* **502,** 64 pp.