Electron. Commun. Probab. **23** (2018), no. 71, 1–9.
https://doi.org/10.1214/18-ECP170
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Fast mixing of metropolis-hastings with unimodal targets

James Johndrow* Aaron Smith†

Abstract

A well-known folklore result in the MCMC community is that the Metropolis-Hastings algorithm mixes quickly for any unimodal target, as long as the tails are not too heavy. Although we've heard this fact stated many times in conversation, we are not aware of any quantitative statement of this result in the literature, and we are not aware of any quick derivation from well-known results. The present paper patches this small gap in the literature, providing a generic bound based on the popular "drift-and-minorization" framework of [19]. Our main contribution is to study two sublevel sets of the Lyapunov function and use path arguments in order to obtain a sharper bound than what can typically be obtained from multistep minorization arguments.

Keywords: Markov chain Monte Carlo; mixing; geometric ergodicity.
AMS MSC 2010: 60J05.
Submitted to ECP on July 28, 2018, final version accepted on September 23, 2018.
Supersedes arXiv:1806.07047.

1 Introduction

The Metropolis algorithm [16] and its generalization, the Metropolis-Hastings algorithm [6], have been exceptionally successful in the numerical approximation of analytically intractable integrals. Because these algorithms are both important and difficult to analyze, there is an enormous literature on the properties of Metropolis-Hastings chains in the statistics, computer science, mathematics and physics communities (see *e.g.* the popular textbooks [17, 14]). Despite the size of this literature, obtaining reasonable quantitative bounds on the convergence rates of specific Markov chains used in statistics can be quite difficult, even when there are good heuristic reasons that convergence should be quick [11, 1]. Recently, the authors needed to use an "obvious" folklore result that does not seem to be in the literature: reasonable Metropolis-Hastings chains targetting unimodal distributions will mix quickly. The main purpose of the paper is to provide a quantitatively useful version of this folklore result (see Theorem 3.1).

We were originally motivated by the need to prove a sharp lower bound on the spectral gap of a Metropolis-Hastings algorithm for logistic regression in a "rare-success" asymptotic regime (see [10]). When the standard deviation of the proposal kernel was similar to the standard deviation of the target distribution, it was straightforward to obtain a quantitatively strong version of the "minorization" condition required by the "drift-and-minorization" approach of [19]. However, this argument becomes much more

*Stanford University, United States of America. E-mail: johndrow@stanford.edu
†University of Ottawa, Canada. E-mail: asmi28@uottawa.ca

delicate when the proposal variance does not closely match the target variance. Although we were motivated by a specific problem, similar problems appear more generally when the proposal kernel of an MCMC algorithm is not perfectly tuned to the target. This sort of (initial) bad tuning can be difficult to avoid in contexts such as [10] where the posterior distribution is very far from Gaussian.

To address this technical problem, we combine pathwise arguments (as studied in [21]) with coupling arguments to obtain reasonable estimates of mixing times inside of compact sublevel sets of the Lyapunov function. We then apply the "drift and minorization" approach of [19] to obtain mixing bounds on the full state space. This argument is presented here for generic random-walk type Metropolis-Hastings, and thus should be broadly useful.

1.1 Related work

Popular approaches for establishing bounds on convergence rates for Markov chains include the Lyapunov-small set techniques of [12, 17, 19], and geometric inequalities such as Poincaré, Cheeger, and log-Sobolev inequalities [2, 4, 3, 13, 20, 21]. Under suitable conditions on the tails of the target and the proposal kernel, drift and minorization arguments show that the Metropolis-Hastings algorithm will converge to the target at an exponential rate [15, 9, 7]. The paper [8] studies essentially the same question addressed in the present paper using Cheeger inequalities, but restricts their attention only to log-concave target distributions.

2 Notation and standing assumptions

Consider a Markov kernel $\mathcal{P}$ with a unique invariant measure $\mu : \mu\mathcal{P} = \mu$. The spectrum of $\mathcal{P}$ is the set S

$$S(\mathcal{P}) = \{\lambda \in \mathbb{C} \setminus \{0\} : (\lambda I - \mathcal{P})^{-1} \text{ is not a bounded linear operator on } L^2(\mu)\},$$

and the spectral gap is defined to be

$$\alpha = 1 - \sup\{|\lambda| \, : \, \lambda \in S, \, \lambda \neq 1\}$$

when the eigenvalue 1 has multiplicity 1, and $\alpha = 0$ else. Define the relaxation time $\tau_{\text{rel}}(\mathcal{P}) \equiv \alpha^{-1}$, and the mixing time τ of $\mathcal{P}$ on a set Θ,

$$\tau \equiv \min\{t \, : \, \sup_{x \in \Theta} \|\delta_x \mathcal{P}^t - \mu\|_{\text{TV}} < 1/4\},$$

which need not be finite.

The following is a strong notion of unimodality on a set:

Definition 2.1. *Fix an interval $[a, b] \subset \mathbb{R}$. Call a function $f : [a, b] \to \mathbb{R}$ unimodal with mode $m \in I$ if f is monotonely increasing on $[a, m]$ and monotonely decreasing on $[m, b]$.*

Definition 2.2. *Fix a closed convex subset $\Theta \subset \mathbb{R}^d$. Call a continuous function $f : \Theta \to \mathbb{R}$ multivariate unimodal if, for all $x \in \Theta$ and $v \in \mathbb{R}^d$, the function $f_{x,v}(s) \equiv f(x + sv)$ is unimodal.*

Note that f will have a (possibly non-unique) maximum M; we call any point m satisfying $f(m) = M$ a mode of f.

Throughout the remainder of the paper, we fix scale $\epsilon > 0$ for typical step sizes of $\mathcal{P}$, in a way that is made concrete in the context of the following assumptions on $\mathcal{P}$:

Assumption 2.3. *The Markov kernel of interest $\mathcal{P}$ is a Metropolis-Hastings kernel with target distribution μ and proposal kernel Q that satisfies*

1. *$Q(x,\cdot)$ has density $q(x,\cdot)$ and μ has density $p(\cdot)$ with respect to Lebesgue measure.*
2. *q is isotropic, i.e. it is of the form $q(x,y) = q(\|x-y\|)$ for some density q that is unimodal.*
3. *For some $\delta_1, c_1, c_2 > 0$, q satisfies*

$$q(x) \geq \delta_1 \mathbf{1}_{\|x\| \leq \epsilon} \tag{2.1}$$
$$q(x) \leq c_1 e^{-\frac{c_2\|x\|}{\epsilon}}.$$

4. *There exist constants $\gamma \in (0,1)$ and $0 \leq K < \infty$ and a Lyapunov function $V : \mathbb{R}^d \to [0,\infty)$ satisfying*

$$(\mathcal{P}V)(x) \leq \gamma V(x) + K. \tag{2.2}$$

The first three assumptions hold for most Metropolis-Hastings proposal kernels used in practice, and we expect them to be easy to verify. The last condition is stronger, and it can be difficult to verify that the condition holds with reasonably small constants γ, K. However, Lyapunov functions do exist under fairly mild conditions that have been well-studied (see *e.g.* [15], [9]).

For chains of this form, define the Metropolis-Hastings acceptance probability by

$$\alpha(x,y) \equiv 1 \wedge \frac{p(y)q(y,x)}{p(x)q(x,y)}.$$

3 Main result

For any $\Theta \subset \mathbb{R}^d$ with $\mu(\Theta) > 0$, denote by μ_Θ and p_Θ the usual restrictions of μ, p to Θ:

$$\mu_\Theta(A) = \frac{\mu(\Theta \cap A)}{\mu(\Theta)}, \quad p_\Theta(x) = \frac{1}{\mu(\Theta)} p(x) \mathbf{1}_{x \in \Theta}.$$

Similarly, denote by $\mathcal{P}_\Theta$ the usual restriction of $\mathcal{P}$ to Θ. That is, $\mathcal{P}_\Theta$ is a Metropolis-Hastings chain with proposal kernel Q and target distribution μ_Θ. Denote by $B_r(x)$ a Euclidean ball of radius r centered at $x \in \mathbb{R}^d$. Our main result is the following

Theorem 3.1. *Let $\mathcal{P}$ be a Metropolis-Hastings transition kernel on $\mathbb{R}$ satisfying the conditions of Assumption 2.3. Let $\Theta \subset \mathbb{R}$ be a set satisfying*

1. *$0 < p < \infty$ is unimodal on Θ with mode $m \in \Theta$, and $\Theta \subseteq B_L(m)$ for some constant $L > 0$.*
2. *p satisfies*

$$\inf_{x \in B_{2\epsilon}(m)} p(x) > \frac{15}{16} p(m), \tag{3.1}$$

   i.e. it is "almost constant" on a ball of radius 2ϵ around the mode.

Then there exists a constant $C = C(c_1, c_2, \delta_1) < \infty$ such that the mixing time τ of $\mathcal{P}_\Theta$ satisfies

$$\tau < C\epsilon^{-3}\delta_1^{-1} L^4 p_\Theta(m). \tag{3.2}$$

If the set Θ is "small but not too small", i.e.

$$\Theta \supset \left\{ x \in \mathbb{R}^d \,:\, V(x) \leq \frac{8}{1-\gamma}\left(\frac{4K}{1-\gamma} + KC\epsilon^{-2}\delta_1^{-1} L^3 p_\Theta(m) \right) \right\}, \tag{3.3}$$

we also have

$$\tau_{\mathrm{rel}}(\mathcal{P}) \leq C\epsilon^{-3}\delta_1^{-1} L^4 p_\Theta(m). \tag{3.4}$$

Although our final result is restricted to $\mathbb{R}$, several of the Lemmas used in proving the result hold with almost no changes in $\mathbb{R}^d$ and could be used to prove similar results for higher-dimensional target distributions. Thus, we prove most of the results in $\mathbb{R}^d$ and specialize to the case of $\mathbb{R}$ for the final Lemma.

4 Proofs

We break the proof up into three lemmas, each of which might be individually useful for proving similar results.

4.1 Mixing: from very small sets to small sets

The first Lemma shows that a Lyapunov condition combined with a bound on the mixing time τ of $\mathcal{P}_\Theta$ for a "small" sublevel set Θ of V allows us to bound the spectral gap by the inverse of the mixing time. The key idea is that a Markov chain started from a point x inside of a "very small" sublevel set is unlikely to escape from the slightly larger "small" set within its first τ steps. This allows us to minorize $\mathcal{P}^\tau$ by μ_Θ inside of the "very small" set $\{x : V(x) < 4K(1-\gamma)^{-1}\}$ and then apply the usual Harris theorem to obtain a bound on the spectral gap. In essence we use the Lyapunov condition and the mixing time bound to obtain a minorization condition on the time scale of the mixing time – that is, for $\mathcal{P}^\tau$ – rather than directly showing a multistep minorization condition.

Lemma 4.1. *Suppose that V is a Lyapunov function of $\mathcal{P}$ satisfying* (2.2)*, $\Theta \subset \mathbb{R}^d$ satisfies* (3.3)*, and the mixing time of $\mathcal{P}_\Theta$ is $\tau < \infty$. Then there exists $C = C(\gamma, K)$ independent of τ and d so that the relaxation time of $\mathcal{P}$ is at most*

$$\tau_{\mathrm{rel}}(\mathcal{P}) \leq C\tau.$$

Proof. Let

$$\begin{aligned}\mathcal{S}(R) &= \{x : V(x) \leq R\},\\ R_1 = \frac{4K}{1-\gamma}, &\quad R_2 = 8\left(\frac{4K}{(1-\gamma)^2} + \frac{K\tau}{1-\gamma}\right)\end{aligned}$$

and fix $x \in \mathcal{S}(R_1)$. Let $\{X_t\}_{t\geq 0}$ be a Markov chain with transition kernel $\mathcal{P}$ and initial state $X_0 = x$. Denote by $\kappa = \inf\{t : X_t \in \Theta^c\}$ the first hitting time of $\Theta^c = \{x \in \mathbb{R}^d : x \notin \Theta\}$. By Inequality (2.2) and Markov's inequality,

$$\begin{aligned}\mathbf{P}[\kappa \leq \tau] &\leq \mathbf{P}[\max_{0\leq k\leq \tau} V(X_k) \geq R_2] \leq \sum_{k=0}^{\tau} \mathbf{P}\left[V(X_t) > R_2\right]\\ &\leq R_2^{-1}\sum_{k=0}^{\tau} \mathbf{E}[V(X_t)]\\ &\leq R_2^{-1}\sum_{k=0}^{\tau}\left(\gamma^k V(x) + K\sum_{j=0}^{k-1}\gamma^j\right)\\ &\leq R_2^{-1}\sum_{k=0}^{\tau}\left(\gamma^k V(x) + K\frac{1-\gamma^{k+1}}{1-\gamma}\right)\\ &\leq R_2^{-1}\left(\frac{R_1}{1-\gamma} + K\frac{(1-\gamma)\tau + \gamma^{\tau+1} - \gamma}{(1-\gamma)^2}\right)\\ &\leq R_2^{-1}\left(\frac{R_1}{1-\gamma} + K\frac{\tau}{1-\gamma}\right) \leq \frac{1}{8}.\end{aligned}$$

Applying this bound, the maximal coupling inequality, and the triangle inequality, we obtain the minorization bound:

$$\sup_{x\in R_1}\|\delta_x\mathcal{P}^\tau-\mu_\Theta\|_{\mathrm{TV}}\le\sup_{x\in R_1}\|\delta_x\mathcal{P}^\tau_\Theta-\mu_\Theta\|_{\mathrm{TV}}+\sup_{x\in R_1}\|\delta_x\mathcal{P}^\tau-\delta_x\mathcal{P}^\tau_\Theta\|_{\mathrm{TV}}$$
$$\le\frac{1}{4}+\frac{1}{8}=\frac{3}{8}.$$

So then $\mathcal{P}^\tau$ satisfies

$$\inf_{x\in\mathcal{S}(R_1)}\delta_x\mathcal{P}^\tau(\cdot)\ge\frac{3}{8}\mu_\Theta(\cdot)$$
$$(\mathcal{P}^\tau V)(x)\le\gamma^\tau V(x)+\frac{K(1-\gamma^\tau)}{1-\gamma}.$$

Applying this minorization bound and the Lyapunov bound (2.2) along with Theorem 1 of [5] to the Markov operator $\mathcal{P}^\tau$ implies that we can bound the geometric convergence rate of convergence $\bar{\alpha}$ of $\mathcal{P}^\tau$ via

$$\begin{aligned}\bar{\alpha}&=\inf_{\alpha_0\in(0,5/8)}\left(1-\frac{5}{8}+\alpha_0\right)\vee\frac{2+\frac{4K}{1-\gamma}\frac{\alpha_0(1-\gamma)}{K(1-\gamma^\tau)}\left(\gamma^\tau+\frac{2K(1-\gamma^\tau)(1-\gamma)}{(1-\gamma)4K}\right)}{2+\frac{4K}{1-\gamma}\frac{\alpha_0(1-\gamma)}{K(1-\gamma^\tau)}}\\&=\inf_{\alpha_0\in(0,5/8)}\left(1-\frac{5}{8}+\alpha_0\right)\vee\frac{2+\frac{4\alpha_0}{(1-\gamma^\tau)}\frac{\gamma^\tau+1}{2}}{2+\frac{4\alpha_0}{(1-\gamma^\tau)}}\\&\le\inf_{\alpha_0\in(0,5/8)}\left(1-\frac{5}{8}+\alpha_0\right)\vee\frac{2+\frac{4\alpha_0}{(1-\gamma)}\frac{\gamma+1}{2}}{2+\frac{4\alpha_0}{(1-\gamma)}}<1,\end{aligned}\tag{4.1}$$

where the last line follows because the second term in the maximum is decreasing in τ. This implies that the geometric convergence rate of $\mathcal{P}$ is at most $\bar{\alpha}^{1/\tau}<1$, and so the $L_2(\mu)$ spectral gap of $\mathcal{P}$ is at least

$$(1-\bar{\alpha}^{1/\tau})$$

by an application of Theorem 2 of [18]. Inspection of inequality (4.1) completes the proof. □

4.2 Mixing for unimodal distributions on compact sets

We now show the first of two Lemmas necessary to prove the mixing time bound inside of Θ. This lemma shows that when started from an initial condition very close to the mode, $\mathcal{P}_\Theta$ will mix rapidly. Our approach is to compare $\mathcal{P}_\Theta$ to a chain with transition kernel $\tilde{\mathcal{P}}_\Theta(x,\cdot)=\mu_\Theta(\cdot)$ for all $x\in\Theta$ - this chain simply takes iid samples from its stationary measure.

We can write the transition densities of $\mathcal{P}$ and $\tilde{\mathcal{P}}$ as:

$$p_\Theta(x,y)=\delta_x(y)\int(1-\alpha_\Theta(x,y))Q(x,dy)+\alpha_\Theta(x,y)q(x,y)$$
$$\tilde{p}_\Theta(x,y)=p_\Theta(y),$$

where

$$\alpha_\Theta(x,y)=\frac{p_\Theta(y)q(y,x)}{p_\Theta(x)q(x,y)}.$$

For $A\subset\Theta$, define the A-restricted mixing time

$$\tau_A(\mathcal{P}_\Theta)=\min\{t:\sup_{x\in A}\|\delta_x\mathcal{P}^t_\Theta-\mu_\Theta\|_{\mathrm{TV}}\le 1/8\}.$$

Lemma 4.2. *Suppose that p satisfies* (3.1) *and Q satisfies* (2.1)*. Then*

$$\tau_{B_\epsilon(m)}(\mathcal{P}_\Theta)\le C\log(16)\epsilon^{-3}\delta_1^{-1}L^{d+3}p_\Theta(m)\frac{\pi^{d/2}3^{d+2}}{\Gamma(d/2+1)},$$

where $0<C<\infty$ is a universal constant.

Proof. Our main tool is Theorem 3.2 of [21]. We recall the definition of the "linear" set of paths Γ, consisting of steps of length ϵ, given in Section 2 of [21]. For fixed $x, y \in \Theta$, define the length $b_{x,y} = \lceil \frac{\|x-y\|}{\epsilon} \rceil$. When $\frac{\|x-y\|}{\epsilon}$ is *not* an integer, set

$$\begin{aligned} \gamma_{x,y}^{(i)} &= x + i\,\epsilon(y-x), \quad 0 \le i < b_{x,y} \\ \gamma_{x,y}^{(b_{x,y})} &= y \\ \gamma_{x,y} &= (\gamma_{x,y}^{(0)}, \ldots, \gamma_{x,y}^{(b_{x,y})}). \end{aligned}$$

Then $\Gamma = \{\gamma_{x,y} : (x,y) \in \Theta\}$ is the collection of all such paths of finite length. We say that a pair $(u,v) \in \Theta \times \Theta$ is an ith edge of the path $\gamma_{x,y}$ iff $u = \gamma_{x,y}^{(i-1)}$ and $v = \gamma_{x,y}^{(i)}$. Let E_i be the collection of the ith edges of all paths $\gamma \in \Gamma$, and put $E = \bigcup_{i \in \mathbb{N}} E_i$. As shown in Section 2 of [21], the set of paths Γ satisfies the regularity conditions of Theorem 3.2 of that paper and, for any $(u,v) \in E$, the associated Jacobian satisfies $J_{x,y}(u,v) = b_{x,y}^d$ (see [21, page 5] for details).

Define $\xi(u,v) = \alpha_\Theta(u,v)q(u,v)p_\Theta(u)$ and for any $\gamma_{x,y} \in \Gamma$ put $\|\gamma_{x,y}\|_0 = \sum_{(u,v)\in\gamma_{x,y}} \xi(u,v)^0 = b_{x,y}$. Notice we can also view the comparison kernel $\tilde{\mathcal{P}}_\Theta$ as a Metropolis-Hastings kernel with acceptance probability $\tilde{\alpha}(x,y) = 1$ and proposal $\tilde{q}(x,y) = p_\Theta(y)$. To use Theorem 3.2 of [21], we must bound the geometric constant

$$\begin{aligned} A_0(\Gamma) &= \operatorname{esssup}_{(u,v)\in E} \left\{ \xi(u,v)^{-1} \sum_{\gamma_{x,y} \ni (u,v)} \|\gamma_{x,y}\|_0 \tilde{p}(x)\tilde{\alpha}(x,y)\tilde{q}(x,y)|J_{x,y}(u,v)| \right\} \\ &= \operatorname{esssup}_{(u,v)\in E} \left\{ \xi(u,v)^{-1} \sum_{\gamma_{x,y} \ni (u,v)} b_{x,y} p_\Theta(x) p_\Theta(y) b_{x,y}^d \right\}. \end{aligned}$$

Bounding below by the uniform proposal on $B_\epsilon(x)$ we have using (2.1) and the volume of a unit ball in $\mathbb{R}^d$

$$\xi(u,v) \ge \left(1 \wedge \frac{p_\Theta(v)}{p_\Theta(u)}\right) \frac{\Gamma(d/2+1)}{\pi^{d/2}} \epsilon^{-d} \delta_1 p_\Theta(u) \ge (p_\Theta(u) \wedge p_\Theta(v)) \frac{\Gamma(d/2+1)}{\pi^{d/2}} \epsilon^{-d} \delta_1$$

which is everywhere positive for any $(u,v) \in \gamma_{x,y}$ by the definition of Γ. Define $b = \max_{x,y} b_{x,y} \le 2L\epsilon^{-1} + 1$, we have

$$\begin{aligned} A_0(\Gamma) &= \operatorname{esssup}_{(u,v)\in E} \left\{ \xi(u,v)^{-1} \sum_{\gamma_{x,y} \ni (u,v)} b_{x,y} p_\Theta(x) p_\Theta(y) b_{x,y}^d \right\} \\ &\le \frac{\pi^{d/2}}{\Gamma(d/2+1)} \delta_1^{-1} \epsilon^d (p_\Theta(u) \wedge p_\Theta(v))^{-1} b^{3+d} p_\Theta(x) p_\Theta(y) \\ &\le \frac{\pi^{d/2}\epsilon^d}{\Gamma(d/2+1)} \delta_1^{-1} b^{3+d} \frac{p_\Theta(x) \wedge p_\Theta(y)}{p_\Theta(u) \wedge p_\Theta(v)} p_\Theta(m) \end{aligned} \tag{4.2}$$

where in the second line we used the fact that there are at most ℓ starting points for paths of length ℓ that contain the edge (u,v) and $\sum_{j=1}^{\ell} j = (\ell^2+\ell)/2 \le \ell^2$. Since p is multivariate unimodal (see Definition 2.2), we note that

$$\frac{p_\Theta(x) \wedge p_\Theta(y)}{p_\Theta(u) \wedge p_\Theta(v)} \le 1$$

for any points $u, v \in \{\gamma_{x,y}^{(0)}, \ldots, \gamma_{x,y}^{(b_{x,y})}\}$ on a linear path from x to y, and therefore

$$A_0(\Gamma) \le \frac{\pi^{d/2}\epsilon^d}{\Gamma(d/2+1)} \delta_1^{-1} b^{3+d} p_\Theta(m).$$

Thus, combining with (4.2), we obtain with $b = 2\epsilon^{-1}L + 1$ the maximum length of a path consisting of steps of size ϵ connecting two points inside a ball of radius L

$$A_0(\Gamma) \leq \frac{\pi^{d/2}}{\Gamma(d/2+1)} 3^{2+d} \epsilon^{-3} L^{d+3} \delta_1^{-1} p_\Theta(m),$$

where we used that $\epsilon \leq L$ so that $2\epsilon^{-1}L + 1 < 3\epsilon^{-1}L$. It follows that the spectral gap α of $\mathcal{P}_\Theta$ satisfies

$$\alpha \geq 3^{-(d+2)} \epsilon^3 \delta_1 L^{-(d+3)} p_\Theta(m)^{-1} \frac{\Gamma(d/2+1)}{\pi^{d/2}}. \tag{4.3}$$

By Inequality (3.1), observe that we can write

$$\delta_x \mathcal{P}_\Theta = \frac{15}{16} \mu_\Theta|_{B_{2\epsilon}(m)} + \frac{1}{16} r_x$$

for some "remainder" measure r_x. Applying Proposition 1.1 of [21] and the bound (4.3) on the spectral gap of $\mathcal{P}_\Theta$, this implies there exists some absolute constant $0 < C < \infty$ such that

$$\begin{aligned} \sup_{x \in B_\epsilon(m)} \|\delta_x \mathcal{P}_\Theta^{t+1} - \mu_\Theta\|_{\mathrm{TV}} &\leq \sup_{x \in B_\epsilon(m)} \|\mu_\Theta|_{B_\epsilon(x)} \mathcal{P}_\Theta^t - \mu_\Theta\|_{\mathrm{TV}} + \frac{1}{16} \\ &\leq \frac{1}{8} \end{aligned} \tag{4.4}$$

for all

$$t > C \log(16) \epsilon^{-3} \delta_1^{-1} L^{d+3} p_\Theta(m) \frac{\pi^{d/2} 3^{d+2}}{\Gamma(d/2+1)}. \qquad \square$$

We prove the last lemma:

Lemma 4.3. *Suppose $\mathcal{P}_\Theta$ is a Metropolis-Hastings kernel on $\Theta \subset \mathbb{R}$ satisfying Assumption 2.3. Then there exists some constant $C = C(\delta_1, c_1, c_2) < \infty$ such that*

$$\tau(\mathcal{P}_\Theta) \leq \tau_{[m-\epsilon, m+\epsilon]}(\mathcal{P}_\Theta) + C \frac{L^2}{\epsilon^2}.$$

Proof. Define the function F by

$$F(x, \Delta, U) = x + \Delta \mathbf{1}\{U < \alpha_\Theta(x, x+\Delta)\}.$$

Note that if $\Delta \sim Q(x, \cdot)$ and $U \sim \mathrm{Uniform}(0,1)$ with $\Delta \perp\!\!\!\perp U$, then

$$F(x, \Delta, U) \sim \mathcal{P}_\Theta(x, \cdot),$$

so F defines a *forward mapping representation* of $\mathcal{P}_\Theta$. Next, let $\Delta_t \overset{iid}{\sim} Q(0, \cdot)$ and $U_t \overset{iid}{\sim} \mathrm{Uniform}(0,1)$. We fix $x \in (m+\epsilon, m+L] \cap \Theta$ and consider a Markov chain (X_t, Y_t, Z_t) on $\mathbb{R}^3$, with initial state (x, x, x) and dynamics defined jointly by

$$\begin{aligned} X_{t+1} &= F(X_t, \Delta_t, U_t) \\ Y_{t+1} &= Y_t + \Delta_t \mathbf{1}(Y_t + \Delta_t \in [-L, L]) \\ Z_{t+1} &= Z_t + \Delta_t. \end{aligned}$$

We focus initially on the properties of (X_t, Y_t). Define

$$\begin{aligned} \tau_{\mathrm{hit}}^x(x) &= \inf\{t \geq 0 \,:\, X_t \in [m-\epsilon, m+\epsilon]\} \\ \tau_{\mathrm{hit}}^y(x) &= \inf\{t \geq 0 \,:\, Y_t \in [m-L, m+\epsilon]\}. \end{aligned}$$

Defining $\tau^* = \tau^x_{\text{hit}}(x) \wedge \tau^y_{\text{hit}}(x)$, it is clear that $m + \epsilon \leq X_t \leq Y_t$ for all $t < \tau^*$. We also have

$$\mathbf{P}[X_{t+1} \in [m, m+\epsilon] \mid X_t > m+\epsilon, t < \tau^*] \geq \mathbf{P}[Y_{t+1} \in [m, m+\epsilon] \mid Y_t > m+\epsilon, t < \tau^*]$$

and

$$\mathbf{P}[Y_{t+1} \in [m, m+\epsilon] \mid Y_{t+1} \in [m-L, m+\epsilon], Y_t > m+\epsilon, t < \tau^*] > \delta \equiv \delta(c_1, c_2) > 0,$$

where the second bound comes from Inequality (2.1). Combining these two bounds, we have shown

$$\mathbf{P}[\tau^x_{\text{hit}}(x) \leq t] \geq \delta \mathbf{P}[\tau^y_{\text{hit}}(x) \leq t].$$

Now we just need a bound on $\tau^y_{\text{hit}}(x)$, which we obtain by comparing Y_t and Z_t. By the Berry-Esseen theorem and the sub-exponential tail bound in Inequality (2.1), there exists a $C_1 = C_1(c_1, c_2, \delta_1) < \infty$ such that, for all $t > C_1 \frac{L^2}{\epsilon^2}$,

$$\mathbf{P}[Z_t < m] > \frac{1}{8}.$$

Using again the fact that the tails of Q are sub-exponential (in the sense of Inequality (2.1)), along with the fact that $Y_t \leq Z_t$ for all $t < \min\{s : Z_s < -L\}$, there exists $C_2 = C_2(c_1, c_2, \delta_1) > 0$ such that

$$\mathbf{P}[\min_{0 \leq s \leq t} Y_s \leq m + \epsilon] \geq C_2 \mathbf{P}[Z_t < m] \geq \frac{C_2}{8} \equiv \eta.$$

This implies

$$\mathbf{P}\left[\tau^y_{\text{hit}}(x) > C_3 \frac{L^2}{\epsilon^2}\right] \leq 1 - \eta$$
$$\mathbf{P}\left[\tau^x_{\text{hit}}(x) \leq C_3 \frac{L^2}{\epsilon^2}\right] \geq \delta\eta;$$

for some constant $C_3 > 0$. Using the strong Markov property, we conclude that for all $T \in \mathbb{N}$

$$\mathbf{P}\left[\tau^x_{\text{hit}}(x) > C_3 T \frac{L^2}{\epsilon^2}\right] \leq (1 - \delta\eta)^T. \tag{4.5}$$

We proved this inequality for $x \in [m + \epsilon, L] \cap \Theta$. By the symmetry of the situation, it is clear that this also holds for $x \in [-L, m - \epsilon] \cap \Theta$, and of course it trivially holds for $x \in [m - \epsilon, m + \epsilon] \cap \Theta$. Thus, Inequality (4.5) holds for all $x \in \Theta$.

Combining Inequality (4.4) and Inequality (4.5) (with the choice $T = (-\log(8))/\log(1 - \delta\eta)$), and setting $t = \tau_{[m-\epsilon, m+\epsilon]}(\mathcal{P}_\Theta) + C_3 T \frac{L^2}{\epsilon^2}$, we have for $x \in \Theta$:

$$\begin{aligned}\|\delta_x \mathcal{P}^t - \mu_\Theta\|_{\text{TV}} &\leq \mathbf{P}[\tau^x_{\text{hit}}(x) < C_3 T \frac{L^2}{\epsilon^2}] + \sup_{x \in [m-\epsilon, m+\epsilon] \cap \Theta} \|\delta_x \mathcal{P}^{\tau_{[m-\epsilon, m+\epsilon]}(\mathcal{P}_\Theta)} - \mu_\Theta\|_{\text{TV}} \\ &\leq \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.\end{aligned}$$

This completes the proof of the lemma. □

4.3 Proof of Theorem 3.1

Theorem 3.1 follows quickly from our three main lemmas. Inequality (3.2) is an immediate consequence of Lemmas 4.2 and 4.3. Inequality (3.4) is an immediate consequence of Inequality (3.2) and Lemma 4.1.

References

[1] Persi Diaconis, Kshitij Khare, and Laurent Saloff-Coste, *Gibbs sampling, exponential families and orthogonal polynomials*, Statistical Science **23** (2008), no. 2, 151–178. MR-2446500

[2] Persi Diaconis and Laurent Saloff-Coste, *Comparison theorems for reversible Markov chains*, The Annals of Applied Probability **3** (1993), no. 3, 696–730. MR-1233621

[3] Persi Diaconis and Laurent Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, The Annals of Applied Probability **6** (1996), no. 3, 695–750. MR-1410112

[4] Persi Diaconis and Laurent Saloff-Coste, *Nash inequalities for finite Markov chains*, Journal of Theoretical Probability **9** (1996), no. 2, 459–510. MR-1385408

[5] Martin Hairer and Jonathan C. Mattingly, *Yet another look at Harris' ergodic theorem for Markov chains*, Seminar on Stochastic Analysis, Random Fields and Applications VI, Springer, 2011, pp. 109–117. MR-2857021

[6] W. Keith Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika **57** (1970), no. 1, 97–109. MR-3363437

[7] Søren F. Jarner and Richard L. Tweedie, *Necessary conditions for geometric and polynomial ergodicity of random-walk-type*, Bernoulli **9** (2003), no. 4, 559–578. MR-1996270

[8] Søren F. Jarner and Wai Kong Yuen, *Conductance bounds on the l^2 convergence rate of Metropolis algorithms on unbounded state spaces*, Advances in Applied Probability **36** (2004), no. 1, 243–266. MR-2035782

[9] Søren F. Jarner and Ernst Hansen, *Geometric ergodicity of Metropolis algorithms*, Stochastic processes and their applications **85** (2000), no. 2, 341–361. MR-1731030

[10] James E. Johndrow, Aaron Smith, Natesh Pillai, and David B. Dunson, *MCMC for imbalanced categorical data*, Journal of the American Statistical Association (in press) (2018).

[11] Galin L. Jones and James P. Hobert, *Honest exploration of intractable probability distributions via Markov chain Monte Carlo*, Statistical Science **16** (2001), no. 4, 312–334. MR-1888447

[12] Rafail Khasminskii, *Stochastic stability of differential equations*, 2 ed., Springer, 2011. MR-2894052

[13] Gregory F. Lawler and Alan D. Sokal, *Bounds on the L_2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality*, Transactions of the American Mathematical Society **309** (1988), no. 2, 557–580. MR-0930082

[14] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer, *Markov chains and mixing times*, American Mathematical Soc., 2009. MR-2466937

[15] Kerrie L. Mengersen and Richard L. Tweedie, *Rates of convergence of the Hastings and Metropolis algorithms*, The Annals of Statistics **24** (1996), no. 1, 101–121. MR-1389882

[16] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller, *Equation of state calculations by fast computing machines*, The Journal of Chemical Physics **21** (1953), no. 6, 1087–1092.

[17] Sean P. Meyn and Richard L. Tweedie, *Markov chains and stochastic stability*, Springer, 1993. MR-1287609

[18] Gareth O. Roberts and Jeffrey S. Rosenthal, *Geometric ergodicity and hybrid Markov chains*, Electronic Communications in Probability **2** (1997), no. 2, 13–25. MR-1448322

[19] Jeffrey S. Rosenthal, *Minorization conditions and convergence rates for Markov chain Monte Carlo*, Journal of the American Statistical Association **90** (1995), no. 430, 558–566. MR-1340509

[20] Alistair Sinclair and Mark Jerrum, *Approximate counting, uniform generation and rapidly mixing Markov chains*, Information and Computation **82** (1989), no. 1, 93–133. MR-1003059

[21] Wai Kong Yuen, *Applications of geometric bounds to the convergence rate of Markov chains on $\mathbb{R}^n$*, Stochastic Processes and their Applications **87** (2000), no. 1, 1–23. MR-1751162

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe `http://www.lockss.org/`
[2]EJMS: Electronic Journal Management System `http://www.vtex.lt/en/ejms.html`
[3]IMS: Institute of Mathematical Statistics `http://www.imstat.org/`
[4]BS: Bernoulli Society `http://www.bernoulli-society.org/`
[5]Project Euclid: `https://projecteuclid.org/`
[6]IMS Open Access Fund: `http://www.imstat.org/publications/open.htm`