On the Containment Condition for Adaptive Markov Chain Monte Carlo Algorithms

by

Yan Bai
Department of Statistics
University of Toronto

and

Gareth O. Roberts
Department of Statistics
University of Warwick

and

Jeffrey S. Rosenthal
Department of Statistics
University of Toronto

Technical Report No. 0806 July 15, 2008

TECHNICAL REPORT SERIES

University of Toronto
Department of Statistics

On the Containment Condition for Adaptive Markov Chain Monte Carlo Algorithms

Yan Bai*, Gareth O. Roberts† and Jeffrey S. Rosenthal‡

July 2008; revised May 2009, Dec. 2009

Abstract

This paper considers ergodicity properties of certain adaptive Markov chain Monte Carlo (MCMC) algorithms for multidimensional target distributions. It was previously shown in [18] that Diminishing Adaptation and Containment imply ergodicity of adaptive MCMC. We derive various sufficient conditions to ensure Containment, and connect the convergence rates of algorithms with the tail properties of the corresponding target distributions. Two examples are given to show that Diminishing Adaptation alone does not imply ergodicity. We also present a Summable Adaptive Condition which, when satisfied, proves ergodicity more easily.

1 Introduction

Markov chain Monte Carlo algorithms are widely used for approximately sampling from complicated probability distributions. However, it is often necessary to tune the scaling and other parameters before the algorithm will converge efficiently. *Adaptive* MCMC algorithms modify their transitions on the fly, in an effort to automatically tune the parameters and improve convergence.

Consider a target distribution $\pi(\cdot)$ defined on the state space $\mathcal{X}$ with respect to some σ-field $\mathcal{B}(\mathcal{X})$ ($\pi(x)$ is also used as the density function). Let $\{P_\gamma : \gamma \in \mathcal{Y}\}$ be the family of transition kernels of time homogeneous Markov chains with the same stationary distribution as π, i.e. $\pi P_\gamma = \pi$ for all $\gamma \in \mathcal{Y}$. *An adaptive MCMC* algorithm $\mathbf{Z} := \{(X_n, \Gamma_n) : n \geq 0\}$ can be regarded as lying in the sample path space $\Omega := (\mathcal{X} \times \mathcal{Y})^\infty$ equipped with a σ-field $\mathcal{F}$. For each initial state $x \in \mathcal{X}$ and initial parameter $\gamma \in \mathcal{Y}$, there is a probability measure $\mathbb{P}_{(x,\gamma)}$ such that the probability of the event $[\mathbf{Z} \in A]$ is well-defined for any set $A \in \mathcal{F}$. There is a filtration $\mathcal{G} := \{\mathcal{G}_n : n \geq 0\}$ such that $\mathbf{Z}$ is adapted to $\mathcal{G}$.

Some adaptive MCMC methods use regeneration times and other somewhat complicated constructions [see 9, 7]. However, see Haario et al. [10] proposed an adaptive Metropolis algorithm attempting to optimise the proposal distribution, and proved that a particular version of this algorithm correctly converges strongly to the target distribution. The algorithm can be viewed as a version of the Robbins-Monro stochastic control algorithm [see 2, 15]. The results were then

*Department of Statistics, University of Toronto, Toronto, ON M5S 3G3, CA. yanbai@utstat.toronto.edu

†Department of Statistics, University of Warwick, Coventry CV4 7AL, UK. gareth.o.roberts@warwick.ac.uk

‡Department of Statistics, University of Toronto, Toronto, ON M5S 3G3, CA. jeff@math.toronto.edu Supported in part by NSERC of Canada.

generalized proving convergence of more general adaptive MCMC algorithms [see 4, 1, 24, 3, 5].

A framework of adaptive MCMC is defined as:

1. Given a initial state $X_0 := x_0 \in \mathcal{X}$ and a kernel P_{Γ_0} with $\Gamma_0 := \gamma_0 \in \mathcal{Y}$. At each iteration $n+1$, X_{n+1} is generated from $P_{\Gamma_n}(X_n, \cdot)$;
2. Γ_{n+1} is obtained from some function of $X_0, \cdots, X_{n+1}$ and $\Gamma_0, \cdots, \Gamma_n$.

For $A \in \mathcal{B}(\mathcal{X})$,

$$\mathbb{P}_{(x_0,\gamma_0)}(X_{n+1} \in A \mid \mathcal{G}_n) = \mathbb{P}_{(x_0,\gamma_0)}(X_{n+1} \in A \mid X_n, \Gamma_n) = P_{\Gamma_n}(X_n, A). \quad (1)$$

In the paper, we study adaptive MCMC with the property Eq. (1). We say that the adaptive MCMC $\mathbf{Z}$ is *ergodic* if for any initial state $x_0 \in \mathcal{X}$ and any kernel index $\gamma_0 \in \mathcal{Y}$, $\left\|\mathbb{P}_{(x_0,\gamma_0)}(X_n \in \cdot) - \pi(\cdot)\right\|_{\mathrm{TV}}$ converges to zero eventually where $\|\mu\|_{\mathrm{TV}} = \sup_{A\in\mathcal{B}(\mathcal{X})} |\mu(A)|$.

Containment is defined as that for any $X_0 = x_0$ and $\Gamma_0 = \gamma_0$, for any $\epsilon > 0$, the stochastic process $\{M_\epsilon(X_n, \Gamma_n) : n \geq 0\}$ is bounded in probability $\mathbb{P}_{(x_0,\gamma_0)}$, i.e. for all $\delta > 0$, there is $N \in \mathbb{N}$ such that $\mathbb{P}_{(x_0,\gamma_0)}(M_\epsilon(X_n, \Gamma_n) \leq N) \geq 1-\delta$ for all $n \in \mathbb{N}$, where $M_\epsilon(x, \gamma) = \inf\{n \geq 1 : \left\|P^n_\gamma(x, \cdot) - \pi(\cdot)\right\|_{\mathrm{TV}} \leq \epsilon\}$ is the "ϵ-convergence time".

Diminishing Adaptation is defined as that for any $X_0 = x_0$ and $\Gamma_0 = \gamma_0$, $\lim_{n\to\infty} D_n = 0$ in probability $\mathbb{P}_{(x_0,\gamma_0)}$ where $D_n = \sup_{x\in\mathcal{X}} \left\|P_{\Gamma_{n+1}}(x, \cdot) - P_{\Gamma_n}(x, \cdot)\right\|_{\mathrm{TV}}$ represents the amount of adaptation performed between iterations n and $n+1$.

Theorem 1 ([18])**.** *Ergodicity of an adaptive MCMC algorithm is implied by Containment and Diminishing Adaptation.*

When designing adaptive algorithms, it is not difficult to ensure that Diminishing Adaptation holds. However, Containment may be more challenging, which raises two questions. First, is Containment really necessary? Second, how can Containment be verified in specific examples? In this paper, we will answer the two questions. In Section 2, two examples are given that explain that 1. Ergodicity holds but neither Containment nor Diminishing Adaptation holds; 2. Diminishing Adaptation alone is not sufficient for ergodicity of adaptive MCMC. Note that Containment alone can not guarantee ergodicity was already discussed in [18, see the "One-Two" version running example]. We also will study *simultaneous geometric ergodicity.* A summable adaptive condition is given which can be used to check ergodicity more easily. Some simple conditions for adaptive Metropolis algorithms implying ergodicity are given. In Section 3, the results are applied to two examples. The proofs of Section 2 are shown in Section 4.

2 Main Results

2.1 Toy Examples

In this section, two examples are given to show that either Diminishing Adaptation or Containment is not necessary for ergodicity of adaptive MCMC, and Diminishing Adaptation alone can not guarantee ergodicity. The state space $\mathcal{X}$ in Example 1 is finite. The kernel index space $\mathcal{Y}$ in Example 2 is finite.

Example 1. *Let the state space $\mathcal{X} = \{1, 2\}$ and the transition kernel*

$$P_\theta = \begin{bmatrix} 1-\theta & \theta \\ \theta & 1-\theta \end{bmatrix}.$$

Obviously, for each $\theta \in (0, 1)$, the stationary distribution is uniform on $\mathcal{X}$.

Proposition 1. *For the target distribution and the family of transition kernels in Example 1, consider a state-independent adaptation: at each time $n \geq 1$ choose the transition kernel index $\theta_{n-1} = \frac{1}{(n+1)^r}$ for some fixed $r > 0$ (P_{θ_0} is the initial kernel). Show that*
(i) *For $r > 0$, Diminishing Adaptation holds but Containment does not;*
(ii) *For $r > 1$, $\mu_0 P_{\theta_0} P_{\theta_1} \cdots P_{\theta_n} \to \mu$ where $\mu_0 = (1, 0)^\top$ and $\mu = (\frac{1+\alpha}{2}, \frac{1-\alpha}{2})^\top$ for some $\alpha \in (0, 1)$;*
(iii) *For $0 < r \leq 1$ and a probability measure μ_0 on $\mathcal{X}$, $\mu_0 P_{\theta_0} P_{\theta_1} \cdots P_{\theta_n} \to \mathrm{Unif}(\mathcal{X})$.*

See the proof in Section 4.1.1.

Remark 1. *The chain in Proposition 1 is a time inhomogeneous Markov chain. It can be suited into the framework of adaptive MCMC. Although very simple, it reflects the complexity of adaptive MCMC to some degree.*
1. *For $r > 1$, the limiting distribution of the chain is not uniform. So it shows that Diminishing Adaptation alone cannot ensure ergodicity.*
2. *For $0 < r \leq 1$, the algorithm is ergodic to a uniform distribution, but Containment does not hold. The reason is that although the "ϵ convergence time" goes to infinity (see Eq. (23)), the distance between the chain and the target is decreasing. See another discussion [5, Section 4].*

Proposition 2. *For the target distribution and the family of transition kernels in Example 1, consider a state-independent adaptation: for $k = 1, 2, \cdots$, at each time $n = 2k - 1$ choose the transition kernel index $\theta_{n-1} = 1/2$, and at each time $n = 2k$ choose the transition kernel index $\theta_{n-1} = 1/n$. Both Diminishing Adaptation and Containment do not hold. The chain converges to the target distribution $\mathrm{Unif}(\mathcal{X})$.*

See the proof in Section 4.1.1.

Example 2. *Let the state space $\mathcal{X} = (0, \infty)$, and the kernel index set $\mathcal{Y} = \{-1, 1\}$. The target density $\pi(x) \propto \frac{\mathbb{I}(x>0)}{1+x^2}$ is a half-Cauchy distribution on the positive part of $\mathbb{R}$. At each time n, run the Metropolis-Hastings algorithm where the proposal value Y_n is generated by*

$$Y_n^{\Gamma_{n-1}} = X_{n-1}^{\Gamma_{n-1}} + Z_n \tag{2}$$

with i.i.d standard normal distribution $\{Z_n\}$, i.e. if $\Gamma_{n-1} = 1$ then $Y_n = X_{n-1} + Z_n$, while if $\Gamma_{n-1} = -1$ then $Y_n = \frac{1}{(1/X_{n-1})+Z_n}$. The adaptation is defined as

$$\Gamma_n = -\Gamma_{n-1}\mathbb{I}(X_n^{\Gamma_{n-1}} < \frac{1}{n}) + \Gamma_{n-1}\mathbb{I}(X_n^{\Gamma_{n-1}} \geq \frac{1}{n}), \tag{3}$$

i.e. we change Γ from 1 to -1 when $X < 1/n$, and change Γ from -1 to 1 when $X > n$, otherwise we do not change Γ.

Proposition 3. *The adaptive chain $\{X_n : n \geq 0\}$ defined in Example 2 does not converge weakly to $\pi(\cdot)$. Containment does not hold.*

See the proof in Section 4.1.2.

2.2 Simultaneous Drift Condition and Summable Adaptive Condition

[18] showed that the *simultaneously strongly aperiodically geometrically ergodic* condition (SSAGE) implies Containment. If there is $C \in \mathcal{B}(\mathcal{X})$, a function $V : \mathcal{X} \to [1, \infty)$, $\delta > 0, \lambda < 1$, and $b < \infty$,

such that $\sup_{x \in C} V(x) < \infty$, and
(i) for each γ, $\exists$ a probability measure $\nu_\gamma(\cdot)$ on C with $P_\gamma(x, \cdot) \geq \delta \nu_\gamma(\cdot)$ for all $x \in C$, and
(ii) $P_\gamma V \leq \lambda V + b\mathbb{I}_C$,
we say that the family $\{P_\gamma : \gamma \in \mathcal{Y}\}$ is SSAGE.

The idea of utilizing SSAGE to check Containment is that SSAGE guarantees there is a uniform quantitative bound of $\left\|P_\gamma^n(x, \cdot) - \pi(\cdot)\right\|_{\mathrm{TV}}$ for all $\gamma \in \mathcal{Y}$. However, SSAGE can be generalized a little. First let us review [23, Theorem 5].

Proposition 4. *Suppose a Markov chain $P(x, dy)$ on the state space $\mathcal{X}$. Let $\{X_n : n \geq 0\}$ and $\{Y_n : n \geq 0\}$ be two realizations of $P(x, dy)$. There are a set $C \subset \mathcal{X}$, $\delta > 0$, some integer $m > 0$, and a probability measure ν_m on $\mathcal{X}$ such that*

$$P^m(x, \cdot) \geq \delta \nu_m(\cdot) \text{ for } x \in C.$$

Suppose further that there exist $0 < \lambda < 1$, $b > 0$, and a function $h : \mathcal{X} \times \mathcal{X} \to [1, \infty)$ such that

$$\mathbb{E}\left[h(X_1, Y_1) \mid X_0 = x, Y_0 = y\right] \leq \lambda h(x, y) + b\mathbb{I}_{C \times C}((x, y)).$$

Let $A := \sup_{(x,y) \in C \times C} \mathbb{E}[h(X_m, Y_m) \mid X_0 = x, Y_0 = y]$, $\mu := \mathcal{L}(X_0)$ be the initial distribution, and π be the stationary distribution. Then for any $j > 0$,

$$\|\mathcal{L}(X_n) - \pi\|_{\mathrm{TV}} \leq (1 - \delta)^{[j/m]} + \lambda^{n - jm + 1} A^{j-1} \mathbb{E}_{\mu \times \pi}[h(X_0, Y_0)].$$

To make use of Proposition 4, we consider the *simultaneously geometrically ergodic* condition (SGE) studied by [22]. If there is $C \in \mathcal{B}(\mathcal{X})$, some integer $m \geq 1$, a function $V : \mathcal{X} \to [1, \infty)$, $\delta > 0$, $\lambda < 1$, and $b < \infty$, such that $\sup_{x \in C} V(x) < \infty$, $\pi(V) < \infty$, and
(i) C is a uniform ν_m-small set, i.e., for each γ, $\exists$ a probability measure $\nu_\gamma(\cdot)$ on C with $P_\gamma^m(x, \cdot) \geq \delta \nu_\gamma(\cdot)$ for all $x \in C$, and
(ii) $P_\gamma V \leq \lambda V + b\mathbb{I}_C$,
we say that the family $\{P_\gamma : \gamma \in \mathcal{Y}\}$ is SGE.

Note that the difference between SGE and SSAGE is that a uniform minorization set C for all P_γ is assumed in SSAGE, however a uniform small set C is assumed in SGE [see the definitions of minorization set and small set in 14, Chapter 5].

Theorem 2. *SGE implies Containment.*

See the proof in Section 4.2.

Corollary 1. *Consider the family $\{P_\gamma : \gamma \in \mathcal{Y}\}$ of Markov chains on $\mathcal{X} \subset \mathbb{R}^d$. Suppose that for any compact set $C \in \mathcal{B}(\mathcal{X})$, there exist some integer $m > 0$, $\delta > 0$ and a measure $\nu_\gamma(\cdot)$ on C for $\gamma \in \mathcal{Y}$ such that $P_\gamma^m(x, \cdot) \geq \delta \nu_\gamma(\cdot)$ for all $x \in C$. Suppose that there is a function $V : \mathcal{X} \to (1, \infty)$ such that for any compact set $C \in \mathcal{B}(\mathcal{X})$, $\sup_{x \in C} V(x) < \infty$, $\pi(V) < \infty$, and*

$$\limsup_{|x| \to \infty} \sup_{\gamma \in \mathcal{Y}} \frac{P_\gamma V(x)}{V(x)} < 1. \tag{4}$$

Then for any adaptive strategy using only $\{P_\gamma : \gamma \in \mathcal{Y}\}$, Containment holds.

See the proof in Section 4.2.

Convergence with sub-geometric rates is studied using a sequence of drift conditions in [25]. It was shown by [12] that if there exist a test function $V \geq 1$, positive constants c and b, a petite set C and $0 \leq \alpha < 1$ such that

$$PV \leq V - cV^{\alpha} + b\mathbb{I}_C, \tag{5}$$

then Markov chain converges to stationary distribution with a polynomial rate. [5] showed that adaptive MCMC of all Markov transition kernel with simultaneous polynomial drift is ergodic under some conditions. The following proposition is a part of the result.

Proposition 5. *Consider an adaptive MCMC algorithm on a state space $\mathcal{X}$. Suppose that there is a set $C \subset \mathcal{X}$ with $\pi(C) > 0$, some constant $\delta > 0$, some integer $m > 0$, and some probability measure $\nu_\gamma(\cdot)$ on $\mathcal{X}$ such that $P_\gamma^m(x, \cdot) \geq \delta\mathbb{I}_C(x)\nu_\gamma(\cdot)$ for $\gamma \in \mathcal{Y}$. Suppose that there are some constants $\alpha \in (0, 1)$, $\beta \in (0, 1]$, $b' > b > 0$, $c > 0$, and some measurable function $V(x) : \mathcal{X} \to [1, \infty)$ with $cV(x) > b'$ on C^c, $\sup_{x \in C} V(x) < \infty$ such that*

$$P_\gamma V \leq V - cV^{\alpha} + b\mathbb{I}_C, \qquad \forall \gamma \in \mathcal{Y}. \tag{6}$$

Then for any adaptive strategy using $\{P_\gamma : \gamma \in \mathcal{Y}\}$ Containment holds.

The idea for the proof is to find the uniform upper bound of $\left\|P_\gamma^n(x, \cdot) - \pi(\cdot)\right\|_{\mathrm{TV}}$. The bound is just dependent of $V(x)$, δ, n, $\pi(V^\beta)$, and C. Since all the transition kernels satisfy the simultaneous polynomial drift condition (Eq. (6)), $\{V(X_n) : n \geq 0\}$ is bounded in probability can be shown. So, Containment holds. [3] study Markovian Adaptation (the joint process $\{(X_n, \Gamma_n) : n \geq 0\}$ is a Markov chain) and give the similar result as the above proposition. But Proposition 5 can be applied to more general adaptive MCMC satisfying Eq. (1) [see details in 5].

In the following result, we use a simple coupling method to show that one summable adaptive condition implies ergodicity of adaptive MCMC.

Proposition 6. *Consider an adaptive MCMC $\{X_n : n \geq 0\}$ on the state space $\mathcal{X}$ with the kernel index space $\mathcal{Y}$. Under the following conditions:*
(i) *$\mathcal{Y}$ is finite. For any $\gamma \in \mathcal{Y}$, P_γ is ergodic with the stationary distribution π;*
(ii) *At each time n, Γ_n is a deterministic measurable function of $X_0, \cdots, X_n, \Gamma_0, \cdots, \Gamma_{n-1}$;*
(iii) *For any initial state $x_0 \in \mathcal{X}$ and any initial kernel index $\gamma_0 \in \mathcal{Y}$,*

$$\sum_{n=1}^{\infty} \mathbb{P}(\Gamma_n \neq \Gamma_{n-1} \mid X_0 = x_0, \Gamma_0 = \gamma_0) < \infty, \tag{7}$$

the adaptive MCMC $\{X_n : n \geq 0\}$ is ergodic with the stationary distribution π.

See the proof in Section 4.2.

Remark 2. *In Example 2, the transition kernel is changed when $X_n^{\Gamma_{n-1}}$ reaches below the bound $1/n$. It can be shown that if the boundary is defined as $1/n^r$ with $r > 1$, the adaptive algorithm is ergodic with half-cauchy distribution because of Proposition 6. To show it, we only need to adopt the procedure in Lemma 2 to check Eq. (7).*

2.3 Adaptive Metropolis algorithm

The target density $\pi(\cdot)$ is defined on the state space $\mathcal{X} \subset \mathbb{R}^d$. In what follows, we shall write $\langle \cdot, \cdot \rangle$ for the usual scalar product on $\mathbb{R}^d$, $|\cdot|$ for the Euclidean and the operator norm, $n(z) := z/|z|$ for the normed vector of z, ∇ for the usual differential (gradient) operator, $m(x) := \nabla\pi(x)/|\nabla\pi(x)|$, $\mathrm{B}^d(x,r) = \{y \in \mathbb{R}^d : |y-x| < r\}$ for the hyperball on $\mathbb{R}^d$ with the center x and the radius r, $\bar{\mathrm{B}}^d(x,r)$ for the closure of the hyperball, and $\mathrm{Vol}(A)$ for the volume of the set $A \subset \mathbb{R}^d$.

Say an adaptive MCMC is an *Adaptive Metropolis-Hastings algorithm* if each kernel P_γ is from a Metropolis-Hastings algorithm

$$P_\gamma(x,dy) = \alpha_\gamma(x,y)Q_\gamma(x,dy) + \left[1 - \int_{\mathcal{X}} \alpha_\gamma(x,z)Q_\gamma(x,dz)\right]\delta_x(dy) \tag{8}$$

where $Q_\gamma(x,dy)$ is the proposal distribution, $\alpha_\gamma(x,y) := \left(\frac{\pi(y)q_\gamma(y,x)}{\pi(x)q_\gamma(x,y)} \wedge 1\right)\mathbb{I}(y \in \mathcal{X})$, and μ_d is Lebesgue measure. Say an adaptive Metropolis-Hastings algorithm is an *Adaptive Metropolis algorithm* if each $q_\gamma(x,y)$ is symmetric, i.e. $q_\gamma(x,y) = q_\gamma(x-y) = q_\gamma(y-x)$.

[11] give conditions which imply geometric ergodicity of symmetric random-walk-based Metropolis algorithm on $\mathbb{R}^d$ for target distribution with lighter-than-exponential tails, [see other related results in 13, 20]. Here, we extend their result a little for target distributions with exponential tails.

Definition 1 (Lighter-than-exponential tail)**.** *The density $\pi(\cdot)$ on $\mathbb{R}^d$ is lighter-than-exponentially tailed if it is positive and has continuous first derivatives such that*

$$\limsup_{|x|\mapsto\infty} \langle n(x), \nabla \log \pi(x)\rangle = -\infty. \tag{9}$$

Remark 3. 1. *The definition implies that for any $r > 0$, there exists $R > 0$ such that*

$$\frac{\pi(x+\alpha n(x)) - \pi(x)}{\pi(x)} \le -\alpha r, \text{ for } |x| \ge R, \alpha > 0.$$

It means that $\pi(x)$ is exponentially decaying along any ray, but with the rate r tending to infinity as x goes to infinity.

2. *The normed gradient $m(x)$ will point towards the origin, while the direction $n(x)$ points away from the origin. For Definition 1, $\langle n(x), \nabla \log \pi(x)\rangle = \frac{|\nabla\pi(x)|}{\pi(x)}\langle n(x), m(x)\rangle$. Even $\limsup_{|x|\mapsto\infty} \langle n(x), m(x)\rangle < 0$, Eq. (9) might not be true. E.g. $\pi(x) \propto \frac{1}{1+x^2}$, $x \in \mathbb{R}$. $m(x) = -n(x)$ so that $\langle n(x), m(x)\rangle = -1$. $\langle n(x), \nabla \log \pi(x)\rangle = -\frac{2|x|}{1+x^2}$ so $\lim_{|x|\mapsto\infty} \langle n(x), \nabla \log \pi(x)\rangle = 0$.*

Definition 2 (Exponential tail)**.** *The density function $\pi(\cdot)$ on $\mathbb{R}^d$ is exponentially tailed if it is a positive, continuously differentiable function on $\mathbb{R}^d$, and*

$$\eta_2 := -\limsup_{|x|\to\infty} \langle n(x), \nabla \log \pi(x)\rangle > 0. \tag{10}$$

Remark 4. *There exists $\beta > 0$ such that for x sufficiently large,*

$$\langle n(x), \nabla \log \pi(x)\rangle = \langle n(x), m(x)\rangle |\nabla \log \pi(x)| \le -\beta.$$

Further, if $0 < -\langle n(x), m(x)\rangle \le 1$, then $|\nabla \log \pi(x)| \ge \beta$.

Define the *symmetric proposal density family* $\mathfrak{C} := \{q : q(x,y) = q(x-y) = q(y-x)\}$. Our ergodicity result for adaptive Metropolis algorithms is based on the following assumptions.

Assumption 1 (Target Regularity)**.** *The target distribution is absolutely continuous w.r.t. Lebesgue measure μ_d with a density π bounded away from zero and infinity on compact sets, and $\sup_{x\in\mathcal{X}} \pi(x) < \infty$.*

Assumption 2 (Target Strongly Decreasing)**.** *The target density π has continuous first derivatives and satisfies*

$$\eta_1 := -\limsup_{|x|\mapsto\infty} \langle n(x), m(x)\rangle > 0. \tag{11}$$

Assumption 3 (Proposal Uniform Local Positivity)**.** *Assume that $\{q_\gamma : \gamma \in \mathcal{Y}\} \subset \mathfrak{C}$. There exist $\zeta > 0$ such that*

$$\iota := \inf_{\gamma\in\mathcal{Y}} \inf_{|z|\leq\zeta} q_\gamma(z) > 0. \tag{12}$$

Given $0 < p < q < \infty$, for $u \in S^{d-1}$ (S^{d-1} is the unit hypersphere in $\mathbb{R}^d$.) and $\theta > 0$, define

$$C_{p,q}(u,\theta) := \left\{ z = a\xi \mid p \leq a \leq q,\, \xi \in S^{d-1},\, |\xi - u| < \theta/3 \right\}. \tag{13}$$

Assumption 4 (Proposal Moment Condition)**.** *Suppose the target density π is exponentially tailed and $\{q_\gamma : \gamma \in \mathcal{Y}\} \subset \mathfrak{C}$. Under Assumptions 2, assume that there are $\epsilon \in (0, \eta_1)$, $\beta \in (0, \eta_2)$, δ, and Δ with $0 < \frac{3}{\beta\epsilon} \leq \delta < \Delta \leq \infty$ such that*

$$\inf_{(u,\gamma)\in S^{d-1}\times\mathcal{Y}} \int_{C_{\delta,\Delta}(u,\epsilon)} |z|\, q_\gamma(z)\mu_d(dz) > \frac{3(e+1)}{\beta\epsilon(e-1)}. \tag{14}$$

Remark 5. *Under Assumption 3, let $\tilde{P}(x,dy)$ be the transition kernel of Metropolis-Hastings algorithm with the proposal distribution $\tilde{Q}(x,\cdot) \sim \mathrm{Unif}(\bar{\mathrm{B}}^d(x,\zeta/2))$. For any $\gamma \in \mathcal{Y}$, $P_\gamma(x,dy) \geq \iota\mathrm{Vol}(\bar{\mathrm{B}}^d(0,\zeta/2))\tilde{P}(x,dy)$. Under Assumptions 1, by [20, Theorem 2.2], any compact set is a small set for $\tilde{P}$ so that any compact set is a uniform small set for all P_γ.*

Remark 6. 1. *Assumption 4 means that the proposal family has uniform lower bound of the first moment on some local cone around the origin. The condition specifies that the tails of all proposal distributions can not be too light, and the quantity of the lower bound is given and dependent on the tail-decaying rate η_2 and the strongly decreasing rate η_1 of target distribution. Assumptions 1-4 are used to check SGE which is just sufficient to Containment.*
2. *If the proposal distribution in $\{q_\gamma : \gamma \in \mathcal{Y}\} \subset \mathfrak{C}$ is a mixture distribution with one fixed part, then Assumption 4 is relatively easy to check, because the integral in Eq. (14) can be estimated by the fixed part distribution. Especially for the lighter-than-exponentially tailed target, Assumption 4 can be reduced for this case. We will give a sufficient condition for Assumption 4 which can be applied to more general case, see Lemma 1.*

Now, we consider a particular class of target densities with tails which are heavier than exponential tails. It was previously shown by [8] that the Metropolis algorithm converges at any polynomial rate when proposal distribution is compact supported and the log density decreases hyperbolically at infinity, $\log \pi(x) \sim -|x|^s$, for $0 < s < 1$, as $|x| \to \infty$.

Definition 3 (Hyperbolic tail). *The density function $\pi(\cdot)$ is twice continuously differentiable, and there exist $0 < m < 1$ and some finite positive constants d_i, D_i, $i = 1, 2$ such that for large enough $|x|$,*
$0 < d_0 |x|^m \le -\log \pi(x) \le D_0 |x|^m$;
$0 < d_1 |x|^{m-1} \le |\nabla \log \pi(x)| \le D_1 |x|^{m-1}$;
$0 < d_2 |x|^{m-2} \le \left|\nabla^2 \log \pi(x)\right| \le D_2 |x|^{m-2}$.

Assumption 5 (Proposal's Uniform Compact Support). *Under Assumption 3, there exists some $M > \zeta$ such that all $q_\gamma(\cdot)$ with $\gamma \in \mathcal{Y}$ are just supported on $\bar{\mathrm{B}}^d(0, M)$.*

Theorem 3. *An adaptive Metropolis algorithm with Diminishing Adaptation is ergodic, under any condition of the following:*
(i). *Target density π is lighter-than-exponentially tailed, and Assumptions 1 - 3;*
(ii). *Target density π is exponentially tailed, and Assumptions 1 - 4;*
(iii). *Target density π is hyperbolically tailed, and Assumptions 1 - 3 and 5.*

3 Applications

Here we discuss two examples. The first one (Example 3) is from [17] where the proposal density is a fixed distribution of two multivariate normal distributions, one with fixed small variance, another using the estimate of empirical covariance matrix from historical information as its variance. It is a slight variant of the famous adaptive Metropolis algorithm of Haario et al. [10]. In the example, the target density has lighter-than-exponential tails. The second (Example 4) concerns with target densities with truly exponential tails.

Example 3. *Consider a d-dimensional target distribution $\pi(\cdot)$ on $\mathbb{R}^d$ satisfying Assumptions 1 - 2. We perform a Metropolis algorithm with proposal distribution given at the n^{th} iteration by $Q_n(x, \cdot) = N(x, (0.1)^2 I_d/d)$ for $n \le 2d$; For $n > 2d$,*

$$Q_n(x, \cdot) = \begin{cases} (1-\theta)N(x, (2.38)^2\Sigma_n/d) + \theta N(x, (0.1)^2 I_d/d), & \Sigma_n \text{ is positive definite,} \\ N(x, (0.1)^2 I_d/d), & \Sigma_n \text{ is not positive definite,} \end{cases} \tag{15}$$

for some fixed $\theta \in (0, 1)$, I_d is $d \times d$ identity matrix, and the empirical covariance matrix

$$\Sigma_n = \frac{1}{n}\left(\sum_{i=0}^{n} X_i X_i^\top - (n+1)\overline{X}_n \overline{X}_n^\top\right), \tag{16}$$

where $\overline{X}_n = \frac{1}{n+1}\sum_{i=0}^{n} X_i$, is the current modified empirical estimate of the covariance structure of the target distribution based on the run so far.

Remark 7. *The fixed part $N(x, (0.1)^2 I_d/d)$ can be replaced by $\mathrm{Unif}(\mathrm{B}^d(x, \tau))$ for some $\tau > 0$. For targets with lighter-than-exponential tails, τ can be an arbitrary positive value, because Assumption 3 holds. For targets with exponential tails, τ is dependent on η_1 and η_2.*

Remark 8. *The proposal $N(x, (2.38)^2\Sigma/d)$ is optimal in a particular large-dimensional context, [see 21, 16]. Thus the proposal $N(x, (2.38)^2\Sigma_n/d)$ is an effort to approximate this.*

Remark 9. *Commonly, the iterative form of Eq. (16) is more useful,*

$$\Sigma_n = \frac{n-1}{n}\Sigma_{n-1} + \frac{1}{n+1}\left(X_n - \bar{X}_{n-1}\right)\left(X_n - \bar{X}_{n-1}\right)^\top. \tag{17}$$

Proposition 7. *Suppose that the target density π is exponentially tailed. Under Assumptions 1-4, $\left|\overline{X}_n - \overline{X}_{n-1}\right|$ and $\|\Sigma_n - \Sigma_{n-1}\|_M$ converge to zero in probability where where $\|\cdot\|_M$ is matrix norm.*

Proof: Note that in the proof of Theorem 3, some test function $V(x) = c\pi^{-s}(x)$ for some $s \in (0,1)$ and some $c > 0$ is found such that SGE holds.

To check Diminishing Adaptation, it is sufficient to check that both $\|\Sigma_n - \Sigma_{n-1}\|_M$ and $\left|\overline{X}_n - \overline{X}_{n-1}\right|$ converge to zero in probability where $\|\cdot\|_M$ is matrix norm.

By some algebras,

$$\begin{aligned}&\Sigma_n - \Sigma_{n-1}\\ =&\frac{1}{n+1}X_nX_n^\top - \frac{1}{n-1}\left(\frac{1}{n}\sum_{i=0}^{n-1}X_iX_i^\top\right) + \frac{2n}{n^2-1}\overline{X}_{n-1}\overline{X}_{n-1}^\top - \frac{1}{n+1}\left(X_n\overline{X}_{n-1}^\top + \overline{X}_{n-1}X_n^\top\right).\end{aligned}$$

Hence,

$$\begin{aligned}&\|\Sigma_n - \Sigma_{n-1}\|_M\\ \leq\;& \tfrac{1}{n+1}\left\|X_nX_n^\top\right\|_M + \tfrac{1}{n-1}\left\|\tfrac{1}{n}\textstyle\sum_{i=0}^{n-1}X_iX_i^\top\right\|_M + \tfrac{2}{n}\left\|\overline{X}_{n-1}\overline{X}_{n-1}^\top\right\|_M +\\ & \tfrac{1}{n+1}\left\|X_n\overline{X}_{n-1}^\top + \overline{X}_{n-1}X_n^\top\right\|_M.\end{aligned} \tag{18}$$

indent To prove $\Sigma_n - \Sigma_{n-1}$ converges to zero in probability, it is sufficient to check that $\left\|X_nX_n^\top\right\|_M$, $\left\|\frac{1}{n}\sum_{i=0}^{n-1}X_iX_i^\top\right\|_M$, $\left\|\overline{X}_{n-1}\overline{X}_{n-1}^\top\right\|_M$ and $\left\|X_n\overline{X}_{n-1}^\top + \overline{X}_{n-1}X_n^\top\right\|_M$ are bounded in probability.

Since $\limsup_{|x|\to\infty}\langle n(x), \nabla\log\pi(x)\rangle < 0$, there exist some $K > 0$ and some $\beta > 0$ such that

$$\sup_{|x|\geq K}\langle n(x), \nabla\log\pi(x)\rangle \leq -\beta.$$

For $|x| \geq K$, $\frac{\log\pi(y)-\log\pi(x)}{(r-1)|x|} \leq -\beta$ where $r > 1$ and $y = rx$, i.e. $\left(\frac{\pi(y)}{\pi(x)}\right)^{-s} \geq e^{s\beta\frac{r-1}{r}|y|}$. Taking $x_0 \in \mathbb{R}^d$ with $|x_0| = K$, $V(x) = c\pi^{-s}(x_0)\left(\frac{\pi(x)}{\pi(x_0)}\right)^{-s} \geq cae^{s\beta\frac{r-1}{r}|x|}$ for $x = rx_0$, $r > 1$, and $a := \inf_{|y|\leq K}\pi^{-s}(y) > 0$, because of Assumption 1. If $r \geq 2$ then $\frac{r-1}{r} \geq 0.5$. Therefore, as $|x|$ is extremely large, $V(x) \geq |x|^2$. We know that $\sup_n \mathbb{E}[V(X_n)] < \infty$ (See Theorem 18 in [18]).

Since $\left\|X_nX_n^\top\right\|_M := \sup_{|u|=1} u^\top X_nX_n^\top u \leq \sup_{|u|=1}|u|^2|X_n|^2 \leq |X_n|^2$, $\left\|X_nX_n^\top\right\|_M$ is bounded in probability.

Obviously,

$$\left\|\frac{1}{n}\sum_{i=0}^{n-1}X_iX_i^\top\right\|_M \leq \frac{1}{n}\sum_{i=0}^{n-1}\left\|X_iX_i^\top\right\|_M.$$

Then, for $K > 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=0}^{n-1}\left\|X_iX_i^\top\right\|_M > K\right) \leq \frac{1}{K}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}\left[\left\|X_iX_i^\top\right\|_M\right] \leq \frac{1}{K}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}\left[|X_i|^2\right] \leq \frac{1}{K}\sup_n\mathbb{E}[V(X_n)].$$

Hence, $\left\|\frac{1}{n}\sum_{i=0}^{n-1}X_iX_i^\top\right\|_M$ is bounded in probability.

$\left|\overline{X}_n\right| \leq \frac{1}{n+1}\sum_{i=0}^{n}|X_i|$. So,

$$\mathbb{P}(\left|\overline{X}_n\right| > K) \leq \frac{1}{K}\frac{1}{n+1}\sum_{i=0}^{n}\mathbb{E}[|X_i|] \leq \frac{1}{K}\sup_n\mathbb{E}[V(X_n)].$$

$\left|\overline{X}_n\right|$ is bounded in probability. Hence, $\left\|\overline{X}_{n-1}\overline{X}_{n-1}^\top\right\|_M$ is bounded in probability.

Finally,

$$\left\|X_n\overline{X}_{n-1}^\top + \overline{X}_{n-1}X_n^\top\right\|_M \leq 2\left|X_n\right|\left|\overline{X}_{n-1}\right|.$$

Therefore, $\left\|X_n\overline{X}_{n-1}^\top + \overline{X}_{n-1}X_n^\top\right\|_M$ is bounded in probability. □

Theorem 4. *Suppose that the target density π in Example 3 is lighter-than-exponentially tailed. The algorithm in Example 3 is ergodic.*

Proof: Obviously, the proposal densities has uniformly lower bound function. By Theorem 3 and Proposition 7, the adaptive Metropolis algorithm is ergodic. □

The following lemma is used to check Assumption 4.

Lemma 1. *Suppose that the target density π is exponentially tailed and the proposal density family $\{q_\gamma : \gamma \in \mathcal{Y}\} \subset \mathfrak{C}$. Suppose further that there is a function $q^-(z) := g(|z|)$, $q^- : \mathbb{R}^d \to \mathbb{R}^+$ and $g : \mathbb{R}^+ \to \mathbb{R}^+$, some constants $M \geq 0$, $\epsilon \in (0, \eta_1)$, $\beta \in (0, \eta_2)$ and $\frac{3}{\beta\epsilon} \vee M < \delta < \Delta$ such that for $|z| \geq M$ with the property that $q_\gamma(z) \geq q^-(z)$ for $\gamma \in \mathcal{Y}$ and*

$$\frac{(d-1)\pi^{\frac{d-1}{2}}}{2\Gamma(\frac{d+1}{2})}\mathrm{Be}_{r^2}\left(\frac{d-1}{2}, \frac{1}{2}\right)\int_\delta^\Delta g(t)t^d dt > \frac{3(e+1)}{\beta\epsilon(e-1)}, \tag{19}$$

where η_1 is defined in Eq. (10), η_2 is defined in Eq. (11), $r := \frac{\epsilon}{18}\sqrt{36-\epsilon^2}$, and the incomplete beta function $\mathrm{Be}_x(t_1, t_2) := \int_0^x t^{t_1-1}(1-t)^{t_2-1}dt$, then Assumption 4 holds.

Proof: For $u \in S^{d-1}$,

$$\int_{C_{\delta,\Delta}(u,\epsilon)} |z|\, g(|z|)\mu_d(dz) = \int_\delta^\Delta g(t)t^d dt \int_{\left\{\xi\in S^{d-1} : |\xi-u|<\epsilon/3\right\}} \omega(d\xi).$$

where $\omega(\cdot)$ denotes the surface measure on S^{d-1}.

By the symmetry of $u \in S^{d-1}$, let $u = e_d := (\underbrace{0, \cdots, 0}_{d-1}, 1)$. So, the projection from the piece $\left\{\xi \in S^{d-1} : |\xi - u| < \epsilon/3\right\}$ of the hypersphere S^{d-1} to the subspace $\mathbb{R}^{d-1}$ generated by the first $d-1$ coordinates is $d-1$ hyperball $\mathrm{B}^{d-1}(0, r)$ with the center 0 and the radius $r = \frac{\epsilon}{18}\sqrt{36-\epsilon^2}$. Define $f(z) = \sqrt{1-(z_1^2 + \cdots + z_{d-1}^2)}$.

$$\begin{aligned}
\omega\left(\left\{\xi \in S^{d-1} : |\xi - u| < \epsilon/3\right\}\right) &= \int_{\mathrm{B}^{d-1}(0,r)} \sqrt{1+|\nabla f|^2}dz_1\cdots dz_{d-1} \\
= \frac{(d-1)\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d+1}{2})}\int_0^r \frac{\rho^{d-2}}{\sqrt{1-\rho^2}}d\rho &= \frac{(d-1)\pi^{\frac{d-1}{2}}}{2\Gamma(\frac{d+1}{2})}\mathrm{Be}_{r^2}\left(\frac{d-1}{2}, \frac{1}{2}\right).
\end{aligned}$$

Hence,

$$\int_{C_{\delta,\Delta}(u,\epsilon)} |z|\, g(|z|)\mu_d(dz) = \frac{(d-1)\pi^{\frac{d-1}{2}}}{2\Gamma(\frac{d+1}{2})}\mathrm{Be}_{r^2}\left(\frac{d-1}{2}, \frac{1}{2}\right)\int_\delta^\Delta g(t)t^d dt. \tag{20}$$

Therefore, the result holds. □

Example 4. *Consider the standard multivariate exponential distribution $\pi(x) = c\exp(-\lambda |x|)$ on $\mathbb{R}^d$ where $\lambda > 0$. We perform a Metropolis algorithm with proposal distribution in the family $\{Q_\gamma(\cdot)\}_{\gamma\in\mathcal{Y}}$ at the n^{th} iteration where*

$$Q_n(x,\cdot) = \begin{cases} \text{Unif}\left(\mathrm{B}^d(x,\Delta)\right), & n \le 2d, \text{ or } \Sigma_n \text{ is nonsingular,} \\ (1-\theta)N(x,(2.38)^2\Sigma_n/d) + \theta\,\text{Unif}\left(\mathrm{B}^d(x,\Delta)\right), & n > 2d, \text{ and } \Sigma_n \text{ is singular,} \end{cases} \tag{21}$$

for $\theta \in (0,1)$, $\text{Unif}\left(\mathrm{B}^d(x,\Delta)\right)$ is a uniform distribution on the hyperball $\mathrm{B}^d(x,\Delta)$ with the center x and the radius Δ, and Σ_n is as defined in Eq. (16). The problem is: how to choose Δ such that the adaptive Metropolis algorithm is ergodic?

Proposition 8. *There exists a large enough $\Delta > 0$ such that the adaptive Metropolis algorithm of Example 4 is ergodic.*

Proof: We compute that $\nabla\pi(x) = -\lambda n(x)\pi(x)$. So, $\langle n(x), \nabla\log\pi(x)\rangle = -\lambda$ and $\langle n(x), m(x)\rangle = -1$. So, the target density is exponentially tailed, and Assumptions 1 and 2 hold. Obviously, each proposal density is locally positive. Now, let us check Assumption 4 by using Lemma 1. Because

$$\text{Vol}(\mathrm{B}^d(x,\Delta)) = \frac{\Delta^d\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2}+1)},$$

the function $g(t)$ defined in Lemma 1 is equal to $\frac{1}{\text{Vol}(\mathrm{B}^d(x,\Delta))}$. η_1 defined in Eq. (10) and η_2 defined in Eq. (11) are respectively λ and 1. Now, fix any $\epsilon \in (0,1)$ and any $\delta \in (\frac{1}{\lambda},\infty)$. The left hand side of Eq. (19) is

$$\frac{(d-1)\pi^{\frac{d-1}{2}}}{2\Gamma(\frac{d+1}{2})}\mathrm{Be}_{r^2}\left(\frac{d-1}{2},\frac{1}{2}\right)\int_\delta^\Delta g(t)t^d dt = \frac{d(d-1)}{2(d+1)\mathrm{Be}(\frac{d+1}{2},1/2)}\cdot\mathrm{Be}_{r^2}\left(\frac{d-1}{2},\frac{1}{2}\right)\cdot\Delta\left(1-\frac{\delta^{d+1}}{\Delta^{d+1}}\right),$$

where $\mathrm{Be}(x,y)$ and $\mathrm{Be}_r(x,y)$ are beta function and incomplete beta function, r is a function of ϵ defined in Lemma 1.

Once fixed ϵ and δ, the first two terms in the right hand side of the above equation is fixed. Then, as Δ goes to infinity, the whole equation tends to infinity. So, there exists a large enough $\Delta > 0$ such that Eq. (19) holds. By Lemma 1, Assumption 4 holds. Then, by Proposition 9, Containment holds. By Proposition 7, Diminishing Adaptation holds. By Theorem 1, the adaptive Metropolis algorithm is ergodic. □

4 Proofs of the Main Results

4.1 Proofs of Section 2.1

4.1.1 Proofs of Example 1

Proof of Proposition 1: Since the adaptation is state-independent, the stationarity is preserved. So, the adaptive MCMC $X_n \sim \delta P_{\theta_0}P_{\theta_1}P_{\theta_2}\cdots P_{\theta_{n-1}}(\cdot)$ for $n \ge 0$ where $\delta := (\delta^{(1)},\delta^{(2)})$ is the initial distribution.

The part (i). Consider $\left\|P_{\theta_{n+1}}(x,\cdot) - P_{\theta_n}(x,\cdot)\right\|_{\text{TV}}$. For any $x \in \mathcal{X}$,

$$\left\|P_{\theta_{n+1}}(x,\cdot) - P_{\theta_n}(x,\cdot)\right\|_{\text{TV}} = |\theta_{n+1} - \theta_n| \to 0.$$

Thus, for $r > 0$ Diminishing Adaptation holds.

By some algebra,

$$\|P_\theta^n(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} = \frac{1}{2}\left|1-2\theta\right|^n. \tag{22}$$

Hence, for any $\epsilon > 0$,

$$M_\epsilon(X_n,\theta_n) \geq \frac{\log(\epsilon) - \log(1/2)}{\log|1-2\theta_n|} \to +\infty \qquad \text{as } n\to\infty. \tag{23}$$

Therefore, the stochastic process $\{M_\epsilon(X_n,\theta_n) : n \geq 0\}$ is not bounded in probability.

The parts (ii) and (iii). Let $\mu_n := \left(\mu_n^{(1)}, \mu_n^{(2)}\right) := \delta P_{\theta_0}\cdots P_{\theta_n}$. So,

$$\mu_{n+1}^{(1)} = \mu_n^{(1)} - \theta_{n+1}\left(\mu_n^{(1)} - \mu_n^{(2)}\right) \text{ and } \mu_{n+1}^{(2)} = \mu_n^{(2)} + \theta_{n+1}\left(\mu_n^{(1)} - \mu_n^{(2)}\right).$$

Hence,

$$\mu_{n+1}^{(1)} - \mu_{n+1}^{(2)} = \left(\delta^{(1)} - \delta^{(2)}\right)\prod_{k=0}^{n+1}(1-2\theta_k).$$

For $r > 1$, $\prod_{k=0}^{n+1}(1-2\theta_k)$ converges to some $\alpha \in (0,1)$ as n goes to infinity. $\mu_{n+1}^{(1)} - \mu_{n+1}^{(2)} \to \left(\delta^{(1)} - \delta^{(2)}\right)\alpha$. For $0 < r \leq 1$, $\mu_{n+1}^{(1)} - \mu_{n+1}^{(2)} \to 0$. Therefore, for $r > 1$ ergodicity to Uniform distribution does not hold, and for $0 < r \leq 1$ ergodicity holds. $\square$

Proof of Proposition 2: From Eq. (22), for $\epsilon > 0$, $M_\epsilon(X_{2k-1},\theta_{2k-1}) \geq \frac{\log(\epsilon)-\log(1/2)}{\log|1-1/k|} \to \infty$ as $k\to\infty$. So, Containment does not hold.
$\left\|P_{\theta_{2k}}(x,\cdot) - P_{\theta_{2k-1}}(x,\cdot)\right\|_{\mathrm{TV}} = \left|\frac{1}{2} - \frac{1}{2k}\right| \to \frac{1}{2}$ as $k\to\infty$. So Diminishing Adaptation does not hold. Let $\delta := (\delta^{(1)}, \delta^{(2)})$ be the initial distribution and $\mu_n := (\mu_n^{(1)}, \mu_n^{(2)}) = \delta P_{\theta_0}\cdots P_{\theta_n}$. $\mu_n^{(1)} - \mu_n^{(2)} = (\delta^{(1)} - \delta^{(2)})2^{-[n/2]-1}\prod_{k=1}^{[(n+1)/2]}\left(1-\frac{1}{2k}\right) \to 0$ as n goes to infinity. So ergodicity holds. $\square$

4.1.2 Proof of Proposition 3

First, we show that Diminishing Adaptation holds.

Lemma 2. *For the adaptive chain $\{X_n : n \geq 0\}$ defined in Example 2, the adaptation is diminishing.*

Proof: For $\gamma = 1$, obviously the proposal density is $q_\gamma(x,y) = \varphi(y-x)$ where $\varphi(\cdot)$ is the density function of standard normal distribution. For $\gamma = -1$, the random variable $1/x + Z_n$ has the density $\varphi(y - 1/x)$ so the random variable $1/(1/x + Z_n)$ has the density $q_\gamma(x,y) = \varphi(1/y - 1/x)/y^2$.

The proposal density

$$q_\gamma(x,y) = \begin{cases} \varphi(y-x) & \gamma = 1 \\ \varphi(1/y - 1/x)/y^2 & \gamma = -1 \end{cases}$$

For $\gamma = 1$, the acceptance rate is $\min\left(1, \frac{\pi(y)q_\gamma(y,x)}{\pi(x)q_\gamma(x,y)}\right)\mathbb{I}(y \in \mathcal{X}) = \frac{1+x^2}{1+y^2}\mathbb{I}(y > 0)$. For $\gamma = -1$, the acceptance rate is $\min\left(1, \frac{\pi(y)q_\gamma(y,x)}{\pi(x)q_\gamma(x,y)}\right)\mathbb{I}(y \in \mathcal{X}) = \min\left(1, \frac{\frac{1}{1+y^2}\varphi(1/x-1/y)/x^2}{\frac{1}{1+x^2}\varphi(1/y-1/x)/y^2}\right)\mathbb{I}(y > 0) =$

$\min\left(1, \frac{1+x^{-2}}{1+y^{-2}}\right)\mathbb{I}(y>0)$.

So for $\gamma \in \mathcal{Y}$, the acceptance rate is

$$\alpha_\gamma(x,y) := \min\left(1, \frac{\pi(y)q_\gamma(y,x)}{\pi(x)q_\gamma(x,y)}\right)\mathbb{I}(y\in\mathcal{X}) = \min\left(1, \frac{1+x^{2\gamma}}{1+y^{2\gamma}}\right)\mathbb{I}(y\in\mathcal{X}). \tag{24}$$

From Eq. (3), $[\Gamma_n \neq \Gamma_{n-1}] = [X_n^{\Gamma_{n-1}} < 1/n]$. Since the joint process $\{(X_n, \Gamma_n) : n \geq 0\}$ is a time inhomogeneous Markov chain,

$$\begin{aligned}\mathbb{P}(\Gamma_n \neq \Gamma_{n-1}) &= \int_{\mathcal{X}\times\mathcal{Y}} \mathbb{P}(X_n^{\Gamma_{n-1}} < 1/n \mid X_{n-1}=x, \Gamma_{n-1}=\gamma)\mathbb{P}(X_{n-1}\in dx, \Gamma_{n-1}\in d\gamma)\\ &= \int_{\mathcal{X}\times\mathcal{Y}} P_\gamma(x, [t>0 : t^\gamma < 1/n])\mathbb{P}(X_{n-1}\in dx, \Gamma_{n-1}\in d\gamma)\\ &= \int_{[x^\gamma \geq 1/(n-1)]} P_\gamma(x, [t>0 : t^\gamma < 1/n])\mathbb{P}(X_{n-1}\in dx, \Gamma_{n-1}\in d\gamma)\end{aligned}$$

where the second equality is from Eq. (1), and the last equality is from $\mathbb{P}(X_n^{\Gamma_n} \geq 1/n) = 1$ implied by Eq. (3).

So for any $(x,\gamma) \in [(t,s)\in\mathcal{X}\times\mathcal{Y} : t^s \geq 1/(n-1)]$,

$$P_\gamma(x, [t>0 : t^\gamma < 1/n]) = \int_0^\infty \mathbb{I}(y^\gamma < 1/n)q_\gamma(x,y)dy = \int_{-x^\gamma}^{-x^\gamma+1/n} \varphi(z)dz.$$

Since $-x^\gamma + 1/(n-1) < 0$,

$$\frac{1}{n}\varphi(-x^\gamma) \leq P_\gamma(x, [t>0 : t^\gamma < 1/n]) \leq \frac{\varphi(0)}{n}. \tag{25}$$

We have that

$$\mathbb{P}(\Gamma_n \neq \Gamma_{n-1}) \leq \frac{1}{\sqrt{2\pi}n}. \tag{26}$$

Therefore, for any $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}} \left\|P_{\Gamma_n}(x,\cdot) - P_{\Gamma_{n-1}}(x,\cdot)\right\|_{\mathrm{TV}} > \epsilon\right) \leq \mathbb{P}(\Gamma_n \neq \Gamma_{n-1}) \to 0.$$

$\square$

From Eq. (24), at the n^{th} iteration, the acceptance rate is $\alpha_{\Gamma_{n-1}}(X_{n-1}, Y_n) = \min\left(1, \frac{1+X_{n-1}^{2\Gamma_{n-1}}}{1+Y_n^{2\Gamma_{n-1}}}\right)\mathbb{I}(Y_n > 0)$. Let us denote $\tilde{Y}_n := Y_n^{\Gamma_{n-1}}$ and $\tilde{X}_n := X_n^{\Gamma_n}$. The acceptance rate is equal to

$$\min\left(1, \frac{1+\tilde{X}_{n-1}^2}{1+\tilde{Y}_n^2}\right)\mathbb{I}(\tilde{Y}_n > 0).$$

From Eq. (3), $X_n^{\Gamma_n} = X_n^{-\Gamma_{n-1}}\mathbb{I}(X_n^{\Gamma_{n-1}} < 1/n) + X_n^{\Gamma_{n-1}}\mathbb{I}(X_n^{\Gamma_{n-1}} \geq 1/n)$. When Y_n is accepted, i.e. $X_n = Y_n$,

$$[\tilde{Y}_n < 1/n] = [X_n^{\Gamma_{n-1}} < 1/n] \text{ and } X_n^{\Gamma_n} = \tilde{Y}_n^{-1}\mathbb{I}(\tilde{Y}_n < 1/n) + \tilde{Y}_n\mathbb{I}(\tilde{Y}_n \geq 1/n).$$

On the other hand, from Eq. (2), the conditional distribution $\tilde{Y}_n \mid \tilde{X}_{n-1}$ is $N(\tilde{X}_{n-1}, 1)$.

From the above discussion, the chain $\tilde{\mathbf{X}} := \{\tilde{X}_n : n \geq 0\}$ can be constructed according to the following procedure. Define the independent random variables $Z_n \overset{\text{iid}}{\sim} N(0,1)$, $U_n \overset{\text{iid}}{\sim} \text{Bernoulli}(0.5)$, and $T_n \overset{\text{iid}}{\sim} \text{Unif}(0,1)$.

Let $\tilde{X}_0 = X_0^{\Gamma_0}$. At each time $n \geq 1$, define the variable

$$\tilde{Y}_n := \tilde{X}_{n-1} - U_n |Z_n| + (1 - U_n) |Z_n| . \tag{27}$$

Clearly, $-U_n |Z_n| + (1-U_n)|Z_n| \overset{\text{d}}{=} N(0,1)$ ($\overset{\text{d}}{=}$ means equal in distribution).

If $T_n < \min\left(1, \frac{1+\tilde{X}_{n-1}^2}{1+\tilde{Y}_n^2}\right) \mathbb{I}(\tilde{Y}_n > 0)$ then

$$\tilde{X}_n = \mathbb{I}(\tilde{Y}_n < 1/n)\tilde{Y}_n^{-1} + \mathbb{I}(\tilde{Y}_n \geq 1/n)\tilde{Y}_n; \tag{28}$$

otherwise $\tilde{X}_n = \tilde{X}_{n-1}$.

Note that:
1. The process $\tilde{\mathbf{X}}$ is a time inhomogeneous Markov chain.
2. $\mathbb{P}(\tilde{X}_n \geq 1/n) = 1$ for $n \geq 1$.
3. At the time n, U_n indicates the proposal direction ($U_n = 0$: try to jump towards infinity; $U_n = 1$: try to jump towards zero). $|Z_n|$ specifies the step size if the proposal value Y_n is accepted. T_n is used to check whether the proposal value Y_n is accepted or not. When $U_n = 1$ and $\tilde{Y}_n > 0$, Eq. (28) is always run.

For two integers $0 \leq s \leq t$ and a process X and a set $A \subset \mathcal{X}$, denote $[X_{s:t} \in A] := [X_s \in A; X_{s+1} \in A; \cdots ; X_t \in A]$ and $s : t := \{s, s+1, \cdots, t\}$. For a value $x \in \mathbb{R}$, denote the largest integer less than x by $[x]$.

In the following proofs for the example, we use the notation in the procedure of constructing the process $\tilde{\mathbf{X}}$.

Lemma 3. *Let* $a = \left(\frac{1}{2} - \frac{7\sqrt{2}}{12\sqrt{\pi}}\right)^{-2}$. *Given* $0 < r < 1$, *for* $[x] > 12^{\frac{1}{1-r}}$

$$\mathbb{P}\left(\exists i \in (k+1) : (k + [x]^{1+r}),\ \tilde{X}_i < x/2 \mid \tilde{X}_k = x\right) \leq \frac{[x]^{1+r}}{\left(\frac{[x]}{2} - \frac{7\sqrt{2}[x]^r}{\sqrt{\pi}}\right)^2} \leq \frac{a}{[x]^{1-r}}.$$

Proof: The process $\tilde{\mathbf{X}}$ is generated through the underlying processes $\{(\tilde{Y}_j, Z_j, U_j, T_j) : j \geq 1\}$ defined in Eq. (27) - Eq. (28). Conditional on $[\tilde{X}_k = x]$, we can construct an auxiliary chain $\mathbf{B} := \{B_j : j \geq k\}$ that behaves like an asymmetric random walk until $\tilde{\mathbf{X}}$ reaches below $x/2$, and $\mathbf{B}$ is always dominated from above by $\tilde{\mathbf{X}}$.

It is defined as that $B_k = \tilde{X}_k$; For $j > k$, if $\tilde{X}_{j-1} < x/2$ then $B_j := \tilde{X}_j$, otherwise
1. If proposing towards zero ($U_j = 1$) then $\mathbf{B}$ also jumps in the same direction with the step size $|Z_j|$ (in this case, the acceptance rate $\min\left(1, \frac{1+\tilde{X}_{j-1}^2}{1+\tilde{Y}_j^2}\right)$ is equal to 1);
2. If proposing towards infinity ($U_j = 0$), then B_j is assigned the value $B_{j-1} + |Z_j|$ (the jumping direction of $\mathbf{B}$ at the time j is same as $\tilde{\mathbf{X}}$) with the acceptance rate $\frac{1+(x/2)^2}{1+(x/2+|Z_j|)^2}$ (independent of $\tilde{X}_{j-1}$), i.e. for $j > k$,

$$B_j := \mathbb{I}(\tilde{X}_{j-1} < x/2)\tilde{X}_j + \mathbb{I}(\tilde{X}_{j-1} \geq x/2)\left(B_{j-1} - I_j(x)\right) \tag{29}$$

where

$$I_j(x) := U_j |Z_j| - (1-U_j) |Z_j| \mathbb{I}\left(T_j < \frac{1+(x/2)^2}{1+(x/2+|Z_j|)^2} \right). \tag{30}$$

Note that

1. $\{Z_j, U_j, T_j : j > k\}$ are independent so $\{I_j(x) : j > k\}$ are independent.
2. When $\tilde{X}_{j-1} > x/2$ and $U_j = 0$ (proposing towards infinity), the acceptance rate $1 > \frac{1+\tilde{X}_{j-1}^2}{1+\tilde{Y}_j^2} \geq \frac{1+(x/2)^2}{1+(x/2+|Z_j|)^2}$, so that $\left[T_j < \frac{1+(x/2)^2}{1+(x/2+|Z_j|)^2}\right] \subset \left[T_j < \frac{1+\tilde{X}_{j-1}^2}{1+\tilde{Y}_j^2}\right]$ which is equivalent to $[B_j - B_{j-1} = |Z_j|] \subset [\tilde{X}_j - \tilde{X}_{j-1} = |Z_j|]$. Therefore, $\mathbf{B}$ is always dominated from above by $\tilde{\mathbf{X}}$.

Conditional on $[\tilde{X}_k = x]$,

$$[\exists i \in (k+1) : (k+[x]^{1+r}),\ \tilde{X}_i < x/2] \subset [\exists i \in (k+1) : (k+[x]^{1+r}),\ B_i < x/2]$$

and for $i \in (k+1) : (k+[x]^{1+r})$,

$$\begin{aligned}
&[B_{k:(i-1)} \geq x/2; B_i < x/2] \\
\subset&[B_k \geq x/2; B_k - \sum_{l=k+1}^{t-1} I_l(x) \geq x/2 \text{ for all } t \in (k+1) : i; B_k - \sum_{l=k+1}^{i} I_l(x) < x/2].
\end{aligned}$$

So,

$$\begin{aligned}
&\mathbb{P}\left(\exists i \in (k+1) : (k+[x]^{1+r}),\ \tilde{X}_i < x/2 \mid \tilde{X}_k = x\right) \\
\leq&\mathbb{P}\left(\exists i \in (k+1) : (k+[x]^{1+r}),\ B_k - \sum_{j=k+1}^{i} I_j(x) < x/2 \mid B_k = x\right) \\
\leq&\mathbb{P}(\max_{l \in 1:[x]^{1+r}} \tilde{S}_l > x/2) \\
=&\mathbb{P}(\max_{l \in 1:q} \tilde{S}_l > q^{1/(1+r)}/2)
\end{aligned}$$

where $\tilde{S}_0 = 0$ and $\tilde{S}_l = \sum_{j=1}^{l} I_{k+j}(x)$ and $q = [x]^{1+r}$. $\{I_j(x) : k < j \leq k+l\}$ and B_k are independent so that the right hand side of the above equation is independent of k.

By some algebra,

$$0 \leq \mathbb{E}[I_i(x)] = \frac{1}{2}\mathbb{E}\left[\frac{|Z_i|^2 (x+|Z_i|)}{1+(x/2+|Z_i|)^2}\right] \leq \frac{2}{x}\mathbb{E}\left[|Z_i|^2 (1+|Z_i|)\right] < \frac{7\sqrt{2}}{\sqrt{\pi}x},$$

$$\mathrm{Var}[I_i(x)] = \frac{1}{2} + \frac{1}{2}\mathbb{E}\left[|Z_i|^2 \frac{1+(x/2)^2}{1+(x/2+|Z_i|)^2}\right] - \frac{1}{4}\left(\mathbb{E}\left[\frac{|Z_i|^2 (x+|Z_i|)}{1+(x/2+|Z_i|)^2}\right]\right)^2 \in [0,1].$$

Let $\mu_l = \mathbb{E}[\tilde{S}_l]$ and $S_l = \tilde{S}_l - \mu_l$ and note that μ_l is increasing as l increases, and $\mu_q \in [0, \frac{7\sqrt{2}q}{\sqrt{\pi}}]$. So $\{S_i : i = 1, \cdots, q\}$ is a Martingale. By Kolmogorov Maximal Inequality,

$$\begin{aligned}
\mathbb{P}(\max_{l \in 1:q} \tilde{S}_l > q^{1/(1+r)}/2) \leq&\mathbb{P}(\max_{l \in 1:q} S_l > q^{1/(1+r)}/2 - \mu_q) \\
\leq&\frac{q\mathrm{Var}[I_k(x)]}{(q^{1/(1+r)}/2 - \mu_q)^2} \\
\leq&\frac{[x]^{1+r}}{\left(\frac{[x]}{2} - \frac{7\sqrt{2}[x]^r}{\sqrt{\pi}}\right)^2} < \frac{a}{[x]^{1-r}}.
\end{aligned}$$

The last second inequality is from $[x] > 12^{\frac{1}{1-r}} > \left(\frac{14\sqrt{2}}{\sqrt{\pi}}\right)^{\frac{1}{1-r}}$ implying $\frac{[x]}{2} > \frac{7\sqrt{2}[x]^r}{\sqrt{\pi}}$. □

Assume that X_n converges weakly to $\pi(\cdot)$. Take some $c > 1$ such that for the set $D = (1/c, c)$, $\pi(D) = 9/10$. Taking a $r \in (0,1)$, there exists $N > 2c \vee 12^{\frac{1}{1-r}} \vee \frac{a}{0.5}^{\frac{1}{1-r}} \vee 2^{1/r}\exp(\frac{1}{0.8\varphi(-c)r})$ (a is defined in Lemma 3) such that for any $n > N+1$, $\mathbb{P}(X_n \in D) > 0.8$. Since $[X_n \in D] = [X_n^{\Gamma_n} \in D]$ and $\mathbf{X}^{\mathbf{\Gamma}} \stackrel{\mathrm{d}}{=} \tilde{\mathbf{X}}$, $\mathbb{P}(\tilde{X}_n \in D) > 0.8$. So, $\mathbb{P}(\tilde{X}_n > \frac{n}{2}) < 0.2$ for $n > N$.

Let $m = \exp(\frac{1}{0.8\varphi(-c)})(n+1)-1$ that implies $m > n$, $m-n < n^{1+r}$ (because $n > 2^{1/r}\exp(\frac{1}{0.8\varphi(-c)r})$), and $\log(\frac{m+1}{n+1}) = \frac{1}{0.8\varphi(-c)}$. Then

$$0.2 > \mathbb{P}(\tilde{X}_m > \frac{n}{2}) \geq \sum_{j=n}^{m-1} \mathbb{P}(\tilde{X}_j \in D; \tilde{Y}_{j+1} < \frac{1}{j+1}; \tilde{X}_{(j+1):m} > \frac{n}{2}). \tag{31}$$

From Eq. (27) and Eq. (28), $[\tilde{Y}_{i+1} < \frac{1}{i+1}] = [\tilde{X}_{i+1} = \frac{1}{\tilde{Y}_{i+1}} > i+1]$ for any $i > 1$. Consider $j \in n:(m-1)$. Since $\tilde{\mathbf{X}}$ is a time inhomogeneous Markov chain,

$$\begin{aligned}
&\mathbb{P}\left(\tilde{X}_j \in D; \tilde{Y}_{j+1} < \frac{1}{j+1}; \tilde{X}_{(j+1):m} > n/2\right) \\
=&\mathbb{P}(\tilde{X}_j \in D)\mathbb{P}\left(\tilde{X}_{j+1} = \tilde{Y}_{j+1} < \frac{1}{j+1} \mid \tilde{X}_j \in D\right) \\
&\mathbb{P}\left(\tilde{X}_{(j+2):m} > \frac{n}{2} \mid \tilde{X}_{j+1} = \frac{1}{\tilde{Y}_{j+1}} > j+1\right) \\
=&\mathbb{P}(\tilde{X}_j \in D)\mathbb{P}\left(\tilde{X}_{j+1} = \frac{1}{\tilde{Y}_{j+1}} > j+1 \mid \tilde{X}_j \in D\right) \\
&\left(1 - \mathbb{P}\left(\tilde{X}_t \leq n/2 \text{ for some } t \in (j+1):m \mid \tilde{X}_{j+1} = \frac{1}{\tilde{Y}_{j+1}} > j+1\right)\right).
\end{aligned}$$

From Eq. (25), for any $x \in D$,

$$\mathbb{P}(\tilde{Y}_{j+1} < \frac{1}{j+1} \mid \tilde{X}_j = x) = P_1(x, \{t \in \mathcal{X} : t < 1/(j+1)\}) \in \left[\frac{\varphi(-c)}{j+1}, \frac{\varphi(0)}{j+1}\right].$$

So,

$$\mathbb{P}(\tilde{Y}_{j+1} < \frac{1}{j+1} \mid \tilde{X}_j \in D) \geq \frac{\varphi(-c)}{j+1}.$$

Hence, for $x > j+1$,

$$\begin{aligned}
&\mathbb{P}\left(\tilde{X}_t \leq n/2 \text{ for some } t \in (j+1):m \mid \tilde{X}_{j+1} = x\right) \\
\leq&\mathbb{P}\left(\tilde{X}_t \leq x/2 \text{ for some } t \in (j+1):m \mid \tilde{X}_{j+1} = x\right) \\
\leq&\mathbb{P}\left(\tilde{X}_t \leq x/2 \text{ for some } t \in (j+1):(j+[x]^{1+r}) \mid \tilde{X}_{j+1} = x\right) \\
\leq&\frac{a}{[x]^{1-r}} \leq \frac{a}{n^{1-r}},
\end{aligned}$$

because of $x/2 > n/2$, $m - n < n^{1+r}$, and Lemma 3. Thus,

$$\mathbb{P}\left(\tilde{X}_t \le n/2 \text{ for some } t \in (j+1):m \mid \tilde{X}_{j+1} = \frac{1}{\tilde{Y}_{j+1}} > j+1\right) \le \frac{a}{n^{1-r}}.$$

Therefore,

$$\begin{aligned}\mathbb{P}(\tilde{X}_m > \frac{n}{2}) \ge& 0.8\varphi(-c)(1 - \frac{a}{n^{1-r}}) \sum_{j=n}^{m-1} \frac{1}{j+1} \\ \ge& 0.8\varphi(-c)(1 - \frac{a}{n^{1-r}}) \log((m+1)/(n+1)) = (1 - \frac{a}{n^{1-r}}) > 0.5.\end{aligned}$$

Contradiction! By Lemma 2, Containment does not hold.

4.2 Proofs of Section 2.2

Proof of Theorem 2: Let $\{X_n^{(\gamma)} : n \ge 0\}$ and $\{X_n^{(\gamma)} : n \ge 0\}$ be two realizations of P_γ for $\gamma \in \mathcal{Y}$. Define $h(x,y) := (V(x) + V(y))/2$. From (ii) of SGE, $\mathbb{E}[h(X_1^{(\gamma)}, Y_1^{(\gamma)}) \mid X_0^{(\gamma)} = x, Y_0^{(\gamma)} = y] \le \lambda h(x,y) + b\mathbb{I}_{C\times C}((x,y))$. It is not difficult to get $P_\gamma^m V(x) \le \lambda^m V(x) + bm$ so $A := \sup_{(x,y)\in C\times C} \mathbb{E}[h(X_m^{(\gamma)}, Y_m^{(\gamma)}) \mid X_0^{(\gamma)} = x, Y_0^{(\gamma)} = y] \le \lambda^m \sup_C V + bm =: B$.

Consider $\mathcal{L}(X_0^{(\gamma)}) = \delta_x$ and $j := \sqrt{n}$. By Proposition 4,

$$\left\|P_\gamma^n(x,\cdot) - \pi(\cdot)\right\|_{\mathrm{TV}} \le (1-\delta)^{[\sqrt{n}/m]} + \lambda^{n-\sqrt{n}m+1} B^{\sqrt{n}-1}(V(x) + \pi(V))/2. \tag{32}$$

Note that the quantitative bound is dependent of x, n, δ, m, C, V and π, and independent of γ. As n goes to infinity, the uniform quantitative bound of all $\left\|P_\gamma^n(x,\cdot) - \pi(\cdot)\right\|_{\mathrm{TV}}$ tends to zero for any $x \in \mathcal{X}$.

Let $\{X_n : n \ge 0\}$ be the adaptive MCMC satisfying SGE. From (ii) of SGE, $\sup_n \mathbb{E}[V(X_n) \mid X_0 = x, \Gamma_0 = \gamma_0] < \infty$ so the process $\{V(X_n) : n \ge 0\}$ is bounded in probability. Therefore, for any $\epsilon > 0$, $\{M_\epsilon(X_n, \Gamma_n) : n \ge 0\}$ is bounded in probability given any $X_0 = x$ and $\Gamma_0 = \gamma_0$. □

Proof of Corollary 1: From Eq. (4), letting $\lambda = \limsup_{|x|\to\infty} \sup_{\gamma\in\mathcal{Y}} \frac{P_\gamma V(x)}{V(x)} < 1$, there exists some positive constant K such that $\sup_{\gamma\in\mathcal{Y}} \frac{P_\gamma V(x)}{V(x)} < \frac{\lambda+1}{2}$ for $|x| > K$. By $V > 1$, $P_\gamma V(x) < \frac{\lambda+1}{2} V(x)$ for $|x| > K$. $P_\gamma V(x) \le \frac{\lambda+1}{2} V(x) + b\mathbb{I}_{\{z\in\mathcal{X}:|z|\le K\}}(x)$ where $b = \sup_{x\in\{z\in\mathcal{X}:|z|\le K\}} V(x)$. □

Proof of Proposition 6: Fix $x_0 \in \mathcal{X}$, $\gamma_0 \in \mathcal{Y}$. By the condition (iii) and the Borel-Cantelli Lemma, $\forall \epsilon > 0$, $\exists N_0(x_0, \gamma_0, \epsilon) > 0$ such that $\forall n > N_0$,

$$\mathbb{P}_{(x_0,\gamma_0)}(\Gamma_n = \Gamma_{n+1} = \cdots) > 1 - \epsilon/2. \tag{33}$$

Construct a new chain $\{\tilde{X}_n : n \ge 0\}$ which satisfies that for $n \le N_0$, $\tilde{X}_n = X_n$, and for $n \ge N_0$, $\tilde{X}_n \sim P_{\Gamma_{N_0}}^{n-N_0}(\tilde{X}_{N_0}, \cdot)$. So, for any $n > N_0$ and any set $A \in \mathcal{B}(\mathcal{X})$, by the condition (ii),

$$\begin{aligned}&\mathbb{P}_{(x_0,\gamma_0)}(X_n \in A, \Gamma_{N_0} = \Gamma_{N_0+1} = \cdots = \Gamma_{n-1}) \\ =& \int_{\mathcal{X}^{N_0} \cap [\gamma_{N_0} = \cdots = \gamma_{n-1}]} P_{\gamma_0}(x_0, dx_1) \cdots P_{\gamma_{N_0-1}}(x_{N_0-1}, dx_{N_0}) P_{\gamma_{N_0}}^{n-N_0}(x_{N_0}, A)\end{aligned}$$

and

$$\mathbb{P}_{(x_0,\gamma_0)}(\tilde{X}_n \in A) = \int_{\mathcal{X}^{N_0}} P_{\gamma_0}(x_0, dx_1) \cdots P_{\gamma_{N_0-1}}(x_{N_0-1}, dx_{N_0}) P^{n-N_0}_{\gamma_{N_0}}(x_{N_0}, A)$$

So,

$$\left| \mathbb{P}_{(x_0,\gamma_0)}(X_n \in A, \Gamma_{N_0} = \cdots = \Gamma_{n-1}) - \mathbb{P}_{(x_0,\gamma_0)}(\tilde{X}_n \in A) \right| \le \epsilon/2.$$

Since the condition (i) holds, suppose that for some $K > 0$, $\mathcal{Y} = \{y_1, \cdots, y_K\}$. Denote $\mu_i(\cdot) = \mathbb{P}_{(x_0,\gamma_0)}(\tilde{X}_{N_0} \in \cdot \mid \Gamma_{N_0} = y_i)$ for $i = 1, \cdots, K$. Because of the condition (ii), for $n > N_0$,

$$\begin{aligned}
&\mathbb{P}_{(x_0,\gamma_0)}(\tilde{X}_n \in A) \\
=&\sum_{i=1}^{K} \mathbb{P}_{(x_0,\gamma_0)}(\tilde{X}_n \in A, \Gamma_{N_0} = y_i) \\
=&\sum_{i=1}^{K} \int_{\mathcal{X}^{N_0} \cap [\gamma_{N_0} = y_i]} P_{\gamma_0}(x_0, dx_1) \cdots P_{\gamma_{N_0-1}}(x_{N_0-1}, dx_{N_0}) P^{n-N_0}_{y_i}(x_{N_0}, A) \\
=&\sum_{i=1}^{K} \mathbb{P}_{(x_0,\gamma_0)}(\Gamma_{N_0} = y_i) \mu_i P^{n-N_0}_{y_i}(A).
\end{aligned}$$

By the condition (i), there exists $N_1(x_0, \gamma_0, \epsilon, N_0) > 0$ such that for $n > N_1$,

$$\sup_{i \in \{1, \cdots, K\}} \left\| \mu_i P^n_{y_i}(\cdot) - \pi(\cdot) \right\|_{\mathrm{TV}} < \epsilon/2.$$

So, for any $n > N_0 + N_1$, any $A \in \mathcal{B}(\mathcal{X})$,

$$\begin{aligned}
&\left| \mathbb{P}_{(x_0,\gamma_0)}(X_n \in A) - \pi(A) \right| \\
\le&\left| \mathbb{P}_{(x_0,\gamma_0)}(X_n \in A) - \mathbb{P}_{(x_0,\gamma_0)}(\tilde{X}_n \in A) \right| + \\
&\left| \mathbb{P}_{(x_0,\gamma_0)}(\tilde{X}_n \in A) - \pi(A) \right| \\
\le&(\epsilon/2 + \epsilon/2) + \epsilon/2 = 3\epsilon/2.
\end{aligned}$$

Therefore, the adaptive MCMC $\{X_n : n \ge 0\}$ is ergodic with the target distribution π. □

4.3 Proof of Theorem 3

Before we show that Theorem 3, we state [11, Lemma 4.2].

Lemma 4. *Let x and z be two distinct points in $\mathbb{R}^d$, and let $\xi = n(x - z)$. If $\langle \xi, m(y) \rangle \ne 0$ for all y on the line from x to z, then z does not belong to $\left\{ y \in \mathbb{R}^d : \pi(y) = \pi(x) \right\}$.*

Consider the test function $V(x) = c\pi^{-s}(x)$ for some $c > 0$ and $s \in (0, 1)$ such that $V(x) \ge 1$. Note that it is not difficult to check that for $s \in (0, 1)$, $\pi(V) < \infty$ by utilizing Definition 2.

By some algebras,

$$\begin{aligned}
P_\gamma V(x)/V(x) = &\int_{A(x)-x} \left(\frac{\pi^s(x)}{\pi^s(x+z)} \right) q_\gamma(z) \mu_d(dz) + \\
&\int_{R(x)-x} \left(1 - \frac{\pi(x+z)}{\pi(x)} + \frac{\pi^{1-s}(x+z)}{\pi^{1-s}(x)} \right) q_\gamma(z) \mu_d(dz),
\end{aligned}$$

where the *acceptance region* $A(x) := \{y \in \mathcal{X} | \pi(y) \geq \pi(x)\}$, and the *potential rejection region* $R(x) := \{y \in \mathcal{X} | \pi(y) < \pi(x)\}$. From [19, Proposition 3], we have $P_\gamma V(x) \leq r(s)V(x)$ where $r(s) := 1 + s(1-s)^{-1+1/s}$.

Proposition 9 (Exponential tail). *Suppose that the target density π is exponentially tailed. Under Assumptions 1-4, Containment holds.*

Proof: Consider $s \in [0, 1/2)$. Under Assumption 4, let

$$\begin{aligned} h(\alpha, s) &= r'(s) + \frac{1}{(1-s)^2} - \\ &\quad \frac{\alpha}{1-s} \inf_{(u,\gamma)\in S^{d-1}\times\mathcal{Y}} \int_{C_{\delta,\Delta}(u,\epsilon)} |z| \left[e^{-\alpha s|z|} - e^{-\alpha(1-s)|z|}\right] q_\gamma(z)\mu_d(dz) \text{ and} \\ H(\alpha, s) &= 1 + \int_0^s h(\alpha, t)dt \end{aligned}$$

where $\epsilon, \beta, \delta, \Delta$, and $C_{\delta,\Delta}(\cdot,\cdot)$ are defined in Assumption 4. So, $H(\beta\epsilon/3, 0) = 1$ and

$$\frac{\partial H(\beta\epsilon/3, 0)}{\partial s} = h(\beta\epsilon/3, 0) \leq e^{-1} + 1 - \frac{\beta\epsilon(1-e^{-1})}{3} \inf_{(u,\gamma)\in S^{d-1}\times\mathcal{Y}} \int_{C_{\delta,\Delta}(u,\epsilon)} |z|\, q_\gamma(z)\mu_d(dz) < 0.$$

Therefore, there exists $s_0 \in (0, 1/2)$ such that $H(\beta\epsilon/3, s_0) < 1$.

Denote $C(x) := x - C_{\delta,\Delta}(n(x), \epsilon)$ and $C^\top(x) := x + C_{\delta,\Delta}(n(x), \epsilon)$. For $|x| \geq 2\Delta$ and $y \in C(x) \cup C^\top(x)$, $|y| \geq |x| - \Delta \geq \Delta$ so $|n(y) - n(x)| < \epsilon/3$.

Since the target density $\pi(\cdot)$ is exponentially tailed and Assumption 2, for sufficiently large $|x| > K_1$ with some $K_1 > 2\Delta$, $\langle n(x), \nabla \log \pi(x)\rangle \leq -\beta$ and $\langle n(x), m(x)\rangle \leq -\epsilon$. Then there exists some $K_2 > K_1$ such that for $|x| \geq K_2$, $\langle n(y), m(y)\rangle \leq -\epsilon$ for $y \in C(x) \cup C^\top(x)$. Thus, $|\nabla \log \pi(y)| = \frac{\langle n(y), \nabla \log \pi(y)\rangle}{\langle n(y), m(y)\rangle} \geq \beta$. Moreover, $y = x \pm a\xi$ for some $\delta \leq a \leq \Delta$ and $\xi \in S^{d-1}$. So,

$$\langle \xi, m(y)\rangle = \langle \xi - n(x), m(y)\rangle + \langle n(x) - n(y), m(y)\rangle + \langle n(y), m(y)\rangle < -\epsilon/3. \tag{34}$$

Hence, by Lemma 4, for $|x| > K_2$,

$$C(x) \cap \left\{y \in \mathbb{R}^d : \pi(y) = \pi(x)\right\} = \emptyset \text{ and } C^\top(x) \cap \left\{y \in \mathbb{R}^d : \pi(y) = \pi(x)\right\} = \emptyset.$$

For $y = x + a\xi \in C^\top(x)$,

$$\begin{aligned} \pi(y) - \pi(x) &= \int_0^a \langle \xi, \nabla\pi(x + t\xi)\rangle\, dt \\ &= \int_0^a \langle \xi, n(\nabla\pi(x + t\xi))\rangle\, |\nabla\pi(x + t\xi)|\, dt \\ &< -\frac{\epsilon}{3} \int_0^a |\nabla\pi(x + t\xi)|\, dt \leq 0 \end{aligned}$$

so that $C^\top(x) \subset R(x)$. Similarly, $C(x) \subset A(x)$.

Consider the test function $V(x) = c\pi^{-s_0}(x)$ for some $c > 0$ such that $V(x) > 1$. By Assumption 1, for any compact set $C \subset \mathbb{R}^d$, $\sup_{x\in C} V(x) < \infty$.

For any sequence $\{x_n : n \geq 0\}$ with $|x_n| \to \infty$, there exists some $N > 0$ such that $n > N$, $|x_n| > K_2$. We have

$$P_\gamma V(x_n)/V(x_n) = \int_{\{C(x_n)-x_n\}\cup\left\{C^\top(x_n)-x_n\right\}} I_{x_n,s_0}(z) q_\gamma(z)\mu_d(dz) + \int_{\{C(x_n)-x_n\}^c\cap\left\{C^\top(x_n)-x_n\right\}^c} I_{x_n,s_0}(z) q_\gamma(z)\mu_d(dz),$$

where

$$I_{x_n,s_0}(z) = \begin{cases} \frac{\pi^{s_0}(x_n)}{\pi^{s_0}(x_n+z)}, & z \in A(x_n) - x_n, \\ 1 - \frac{\pi(x_n+z)}{\pi(x_n)} + \frac{\pi^{1-s_0}(x_n+z)}{\pi^{1-s_0}(x_n)}, & z \in R(x_n) - x_n. \end{cases}$$

For $z = a\xi \in C^\top(x_n) - x_n$ and $t \in (0, |z|)$, by Eq. (34)

$$\langle \xi, \nabla \log \pi(x_n + t\xi)\rangle = \langle \xi, m(x_n + t\xi)\rangle \left|\nabla \log \pi(x_n + t\xi)\right| < -\epsilon\beta/3.$$

So, by Assumption 4,

$$\frac{\pi(x_n+z)}{\pi(x_n)} = e^{\log \pi(x_n+z) - \log \pi(x_n)} = e^{\int_0^{|z|} \langle \xi, \nabla \log \pi(x_n+t\xi)\rangle dt} \leq e^{-\beta\epsilon|z|/3} \leq e^{-\beta\epsilon\delta/3} \leq e^{-1}.$$

Similarly, for $z = -a\xi \in C(x_n) - x_n$,

$$\frac{\pi(x_n)}{\pi(x_n+z)} \leq e^{-\beta\epsilon|z|/3} \leq e^{-1}.$$

$t^{1-s_0} - t \leq \frac{1}{1-s_0} t^{1-s_0} - t$. Since $t \to \frac{1}{1-s_0} t^{1-s_0} - t$ is an increasing function on $[0, 1]$,

$$\begin{aligned} &\int_{\{C(x_n)-x_n\}\cup\left\{C^\top(x_n)-x_n\right\}} I_{x_n,s_0}(z) q_\gamma(z)\mu_d(dz) \\ \leq &\int_{C(x_n)-x_n} \frac{1}{1-s_0} e^{-s_0\beta\epsilon|z|/3} q_\gamma(z)\mu_d(dz) + \\ &\int_{C^\top(x_n)-x_n} \left(1 - e^{-\beta\epsilon|z|/3} + \frac{1}{1-s_0} e^{-(1-s_0)\beta\epsilon|z|/3}\right) q_\gamma(z)\mu_d(dz). \end{aligned}$$

On the other hand,

$$\begin{aligned} &\int_{\{C(x_n)-x_n\}^c\cap\left\{C^\top(x_n)-x_n\right\}^c} I_{x_n,s_0}(z) q_\gamma(z)\mu_d(dz) \\ \leq & r(s_0) Q_\gamma\left(\{C(x_n) - x_n\}^c \cap \left\{C^\top(x_n) - x_n\right\}^c\right). \end{aligned}$$

Define $K_{x,\gamma}(t) := \int_{C(x)-x} e^{-t|z|} q_\gamma(z)\mu_d(dz) = \int_{C^\top(x)-x} e^{-t|z|} q_\gamma(z)\mu_d(dz)$, and

$$H_{x,\gamma}(\theta, t) := \frac{K_{x,\gamma}(t\theta)}{1-t} + K_{x,\gamma}(0) - K_{x,\gamma}(\theta) + \frac{K_{x,\gamma}((1-t)\theta)}{1-t} + r(t)(1 - 2K_{x,\gamma}(0)).$$

So,

$$P_\gamma V(x_n)/V(x_n) \leq H_{x_n,\gamma}(\beta\epsilon/3, s_0).$$

Clearly, $K_{x,\gamma}(t) \le 1/2$. For $0 \le t < 1/2$,

$$\begin{aligned}
&\frac{\partial H_{x,\gamma}(\theta,t)}{\partial t}\\
=&r'(t)(1-2K_{x,\gamma}(0)) + \frac{K_{x,\gamma}(\theta t)+K_{x,\gamma}(\theta(1-t))}{(1-t)^2} + \frac{\theta}{1-t}\left(K'_{x,\gamma}(\theta t) - K'_{x,\gamma}(\theta(1-t))\right)\\
\le& r'(t) + \frac{1}{(1-t)^2} - \frac{\theta}{1-t}\int_{C(x)-x}\left(e^{-\theta t|z|} - e^{-\theta(1-t)|z|}\right)|z|\, q_\gamma(z)\mu_d(dz)\\
\le& h(\theta,t).
\end{aligned}$$

Since $H_{x,\gamma}(\theta,0)=1$, $H_{x,\gamma}(\theta,t) \le H(\theta,t)$ for $0 \le t < 1/2$. Thus, $H_{x_n,\gamma}(\beta\epsilon/3, s_0) \le H(\beta\epsilon/3, s_0) < 1$ so $\limsup\limits_{|x|\to\infty} \sup\limits_{\gamma\in\mathcal{Y}} \frac{P_\gamma V(x)}{V(x)} < 1$. By Corollary 1, Containment holds. □

Proof of Theorem 3: For (ii), by Proposition 9, Containment holds. Then ergodicity is implied by Containment and Diminishing Adaptation.

For (i), From Assumption 3, for any $\epsilon \in (0,\eta_1)$ and any $u \in S^{d-1}$,

$$\int_{C_{\zeta/2,\zeta}(u,\epsilon)} |z|\, q_\gamma(z)\mu_d(dz) \ge \frac{\iota\zeta \mathrm{Vol}(C_{\zeta/2,\zeta}(u,\epsilon))}{2}$$

where ι is defined in Eq. (12), ζ is defined in Assumption 3, $C_{a,b}(\cdot,\cdot)$ is defined in Eq. (13). The right hand side of the above equation is positive and independent of γ and u. Since target density is lighter-than-exponentially tailed, $\eta_2 := -\limsup_{|x|\to\infty}\langle n(x), \nabla\log\pi(x)\rangle = +\infty$ such that there is some sufficiently large β such that Eq. (14) holds. So, Assumption 4 is satisfied.

For (iii), adopting the proof of [8, Theorem 5], we will show that the simultaneous drift condition Eq. (6) holds. Denote

$$R(g,x,y) := g(y) - g(x) - \langle \nabla g(x), y-x\rangle.$$

Consider the test function $V(x) := 1 + f^s(x)$ where $f(x) := -\log\pi(x)$ for $\frac{2}{m} - 1 < s < \min(\frac{2}{m}, \frac{3}{m} - 2)$ where m is defined in Definition 3.

So,

$$P_\gamma V(x) - V(x) = P_\gamma f^s(x) - f^s(x) = \sum_{j=0}^{4} I_j(x,\gamma),$$

where M is defined in Assumption 5 and

$$\begin{aligned}
I_0(x,\gamma) :=& -s f^{s-1}(x)\,|\nabla f(x)|^2 \int_{R(x)-x\cap\{|z|\le M\}} \langle m(x), n(z)\rangle^2 |z|^2 q_\gamma(z)\mu_d(dz),\\
I_1(x,\gamma) :=& \int_{\{|z|\le M\}} R(f^s, x, x+z) q_\gamma(z)\mu_d(dz),\\
I_2(x,\gamma) :=& \int_{R(x)-x\cap\{|z|\le M\}} R(f^s, x, x+z)\frac{R(\pi,x,x+z)}{\pi(x)} q_\gamma(z)\mu_d(dz)\\
I_3(x,\gamma) :=& \int_{R(x)-x\cap\{|z|\le M\}} R(f^s, x, x+z)\langle \nabla f(x), z\rangle q_\gamma(z)\mu_d(dz)\\
I_4(x,\gamma) :=& \int_{R(x)-x\cap\{|z|\le M\}} \frac{R(\pi,x,x+z)}{\pi(x)}\langle \nabla f^s(x), z\rangle q_\gamma(z)\mu_d(dz).
\end{aligned}$$

By [8, Lemma B.4] and Assumption 5,

$$
\begin{aligned}
|I_1(x,\gamma)| = \mathrm{O}(|x|^{ms-2}),\ |I_2(x,\gamma)| = \mathrm{O}(|x|^{m(s+2)-4}),\\
|I_3(x,\gamma)| = \mathrm{O}(|x|^{m(s+1)-3}),\ |I_4(x,\gamma)| = \mathrm{O}(|x|^{m(s+2)-3}).
\end{aligned}
$$

Note that the $\mathrm{O}(\cdot)$s in the above equations are independent of γ. Since $\frac{2}{m}-1 < s < \min(\frac{2}{m}, \frac{3}{m}-2)$, $|I_1(x,\gamma)|$, $|I_2(x,\gamma)|$, $|I_3(x,\gamma)|$ and $|I_4(x,\gamma)|$ converge to zero as $|x| \to \infty$.

By Assumption 2, for $\epsilon \in (0,\eta_1)$ (η_1 is defined in Eq. (11)), $\langle n(x), m(x)\rangle < -\epsilon$ as $|x|$ is sufficiently large. By Assumption 3, for sufficiently large $|x|$, for any $z \in C_{0,\zeta}(n(x),\epsilon)$ (ζ is defined in Assumption 3, ι is defined in Eq. (12), and $C_{\cdot,\cdot}(\cdot,\cdot)$ is defined in Eq. (13)),

$$
-1 \le \langle m(x), n(z)\rangle = \langle m(x), n(x)\rangle + \langle m(x), n(z)-n(x)\rangle \le -\epsilon + \epsilon/3.
$$

Thus,

$$
\begin{aligned}
I_0(x,\gamma) \le & -\frac{4\epsilon^2 \iota s f^{s-1}(x)\,|\nabla f(x)|^2}{9} \int_{C_{0,\zeta}(n(x),\epsilon)} |z|^2 \mu_d(dz)\\
= & -c_1 f^{s-1}(x)\,|\nabla f(x)|^2 \le c_2 f^{s-(2-m)/m}(x),
\end{aligned}
$$

for some $c_1 > 0$ (independent of x) where $C_{0,\zeta}(n(x),\epsilon) = C_{0,\zeta}(u,\epsilon)$ for any $u \in S^{d-1}$.

So, there exist some $K > 0$ and some $c_3 > 0$ such that $V(x) > 1.1$ and $P_\gamma V(x) - V(x) \le -c_3 V^\alpha(x)$ for $|x| > K$, some $\alpha \in (0,1)$. Let $\tilde{V}(x) := V(x)\mathbb{I}(|x|>K) + \mathbb{I}(|x| \le K)$. So,

$$
P_\gamma \tilde{V}(x) - \tilde{V}(x) \le -c_3 \tilde{V}^\alpha(x) + c_3 \mathbb{I}(|x| \le K).
$$

By Proposition 5, Containment holds. □

5 Conclusions and Discussion

For adaptive Metropolis algorithms (see similar results for adaptive Metropolis-within-Gibbs algorithms in [6]), we provide some conditions only related to properties of the target density and the proposal family. For targets with lighter-than-exponential tails, ergodicity of adaptive Metropolis algorithms can be implied by the uniform local positivity of the family of proposal densities. For targets with exponential tails, ergodicity of adaptive Metropolis algorithms can be implied by both the uniform local positivity and the uniform lower bound of the first moment of the family of proposals.

Recently, there also is some results about this topic, see [24]. They show that if the target density is regular, strongly decreasing, and strongly lighter-than-exponentially tailed ($\limsup_{|x|\to\infty} \frac{\langle n(x), \nabla \log \pi(x)\rangle}{|x|^{\rho-1}} = -\infty$ for some $\rho > 1$) which is used to keep the convexity of outside manifold contour of target densities, then strong law of large number (SLLN) for symmetric random-walk based adaptive Metropolis algorithms holds. Compared with the results, although the conditions do not require that the target density is strongly lighter-than-exponentially tailed, one restriction on proposal density is needed.

[11] show that if under Assumption 2 target density is lighter-than-exponential tailed then random-walk-based Metropolis algorithms are geometrically ergodic. The technique in Proposition 9 can be also applied to MCMC. So, even if target density is exponentially tailed under some moment condition similar as Eq. (14), any random-walk-based Metropolis algorithm is still geometrically ergodic. Careful readers may mention that our symmetry assumption ($q(x,y) = q(x-y) = q(y-x)$) is a little different from the assumption ($q(x,y) = q(|x-y|)$) of [11].

Acknowledgements

We thank Matti Vihola for helpful comments.

References

[1] C Andrieu and E Moulines. On the ergodicity properties of some adaptive Markov Chain Monte Carlo algorithms. . *Ann. Appl. Probab.*, 16(3):1462–1505, 2006.

[2] C. Andrieu and C.P. Robert. Controlled MCMC for optimal sampling. . *Preprint*, 2001.

[3] Y.F. Atchadé and G. Fort. Limit Theorems for some adaptive MCMC algorithms with subgeometric kernels. *Preprint*, 2008.

[4] Y.F. Atchadé and J.S. Rosenthal. On Adaptive Markov Chain Monte Carlo Algorithms. *Bernoulli*, 11(5):815–828, 2005.

[5] Y. Bai. Simultaneous drift conditions on adaptive Markov Chain Monte Carlo algorithms. *http://www.utstat.utoronto.ca/ yanbai/yanbai2.pdf*, 2009a.

[6] Y. Bai. Convergence of Adaptive Markov Chain Monte Carlo Methods. *PhD thesis, Department of Statistics, University of Toronto*, 2009.

[7] A.E. Brockwell and J.B. Kadane. Indentification of regeneration times in MCMC simulation, with application to adaptive schemes. *J. Comp. Graph. Stat*, 14:436–458, 2005.

[8] G. Fort and E. Moulines. V-Subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statist. Prob. Lett.*, 49:401–410, 2000.

[9] W.R. Gilks, G.O. Roberts, and S.K. Sahu. Adaptive Markov chain Monte Carlo. *J. Amer. Statist. Assoc.*, 93:1045–1054, 1998.

[10] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7: 223–242, 2001.

[11] S.F. Jarner and E. Hansen. Geometric ergodicity of Metropolis algorithms. *Stoch. Process. Appl.*, 85:341–361, 2000.

[12] S.F. Jarner and G.O. Roberts. Polynomial convergence rates of Markov Chains. *Ann. Appl. Probab.*, 12(1):224–247, 2002.

[13] K.L. Mengersen and R.L. Tweedie. Rate of convergences of the Hasting and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996.

[14] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* London: Springer-Verlag, 1993.

[15] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Stat.*, 22:400–407, 1951.

[16] G.O. Roberts and J.S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Stat. Sci.*, 16:351–367, 2001.

[17] G.O. Roberts and J.S. Rosenthal. Examples of Adaptive MCMC. *J. Comp. Graph. Stat.*, 2009.

[18] G.O. Roberts and J.S. Rosenthal. Coupling and Ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Prob.*, 44:458–475, 2007.

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

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

[21] G.O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Prob.*, 7:110–120, 1997.

[22] G.O. Roberts, J.S. Rosenthal, and P.O. Schwartz. Convergence propperties of perturbed Markov chains. *J. Appl. Prob.*, 35:1–11, 1998.

[23] J.S. Rosenthal. Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *J. Amer. Stats. Assoc.*, 90:558–566, 1995.

[24] E. Saksman and M. Vihola. On the Ergodicity of the Adaptive Metropolis Algorithms on Unbounded Domains. *Preprint*, 2008.

[25] P. Tuominen and R.L. Tweedie. Subgeometric rates of convergence of f-ergodic Markov chains. *Adv. Appl. Probab.*, 26(3):775–798, 1994.