On the Containment Condition for Adaptive Markov Chain Monte Carlo Algorithms

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

July 2008; last revised January 2009

Abstract

This paper considers ergodicity properties of certain adaptive Markov chain Monte Carlo (MCMC) algorithms for multidimensional target distributions, in particular Adaptive Metropolis and Adaptive Metropolis-within-Gibbs. It was previously shown (Roberts and Rosenthal [21]) 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. An example is 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.

Some adaptive MCMC methods use regeneration times and other somewhat complicated constructions, see [10] and [5]. However, Haario *et al.* [11] 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] and [16]. The results were then generalized proving convergence of more general adaptive MCMC algorithms, see [3] and [1].

It was proved by Roberts and Rosenthal (RR) [21] that Diminishing Adaptation and Containment imply that adaptive MCMC converges to the target distribution. When designing the algorithm, 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. RR prove that an adaptive MCMC satisfying Diminishing Adaptation satisfies Containment if the family $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ is *simultaneously strongly aperiodically geometrically ergodic*, but this may be difficult to check in practice. In this paper, we give some simpler criteria related to proposals to check Containment, more easily.

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

†Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK. g.o.roberts@lancaster.ac.uk

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

After introducing our notation and terminology in Section 2, we present a counter example in Section 3, which demonstrates that Diminishing Adaptation alone is not sufficient for the ergodicity of adaptive MCMC. However, we show in Section 4 that a stronger version of the Diminishing Adaptation alone implies ergodicity of adaptive algorithm. We then give some results which ensure ergodicity for certain adaptive Metropolis algorithms in Section 5 and adaptive Metropolis-within-Gibbs algorithms in Section 6.

2 Preliminaries

We let $\pi(\cdot)$ be a fixed 'target' probability distribution. on a state space $\mathcal{X}$ with σ-field $\mathcal{F}$. The goal of MCMC is to approximately sample from $\pi(\cdot)$ through the use of Markov chains, particularly when $\pi(\cdot)$ is too complicated and multidimensional to facilitate more direct sampling.

We let $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ be a family of Markov chain kernels on $\mathcal{X}$, each of which has $\pi(\cdot)$ as the unique stationary distribution, i.e. $\pi P_\gamma(\cdot) = \pi(\cdot)$ for all $\gamma \in \mathcal{Y}$.

Assuming that P_γ is ϕ-irreducible and aperiodic, this implies that P_γ is *ergodic* for $\pi(\cdot)$, i.e. $\lim_{n\to\infty} \left\|P_\gamma^n(\cdot) - \pi(\cdot)\right\| = 0$, for all x, where $\|\mu(\cdot) - \nu(\cdot)\| = \sup_{A\in\mathcal{F}} |\mu(A) - \nu(A)|$ is the total variation norm. So, if γ is fixed, we know that P_γ will eventually converge to $\pi(\cdot)$.

However, some choices of γ may lead to far less efficient algorithms than others, and it may be difficult to know in advance which choices of γ are preferable. To deal with this, *adaptive MCMC* proposes that at each time n we let the choice of γ be given by a $\mathcal{Y}$-valued random variable Γ_n, updated according to specified rules.

Formally, for $n = 0, 1, 2, \ldots$, we have an $\mathcal{Y}$-valued random variable Γ_n, representing the choice of transition kernel to be used when updating from X_n to X_{n+1}. We let

$$\mathcal{G}_n := \sigma\left(X_0, \ldots, X_n, \Gamma_0, \ldots, \Gamma_n\right)$$

be the filtration generated by $\{(X_n, \Gamma_n)\}$. Thus,

$$\mathbf{P}\left[X_{n+1} \in B | X_n = x, \Gamma_n = \gamma, \mathcal{G}_{n-1}\right] = P_\gamma(x, B), \qquad x \in \mathcal{X}, \gamma \in \mathcal{Y}, B \in \mathcal{F}, \tag{1}$$

while the conditional distribution of Γ_{n+1} given $\mathcal{G}_n$ is to be specified by the particular adaptive algorithm being used. We let

$$A^{(n)}((x,\gamma), B) = \mathbf{P}[X_n \in B | X_0 = x, \Gamma_0 = \gamma], \qquad B \in \mathcal{F},$$

record the conditional probabilities for X_n for the adaptive algorithm, given the initial conditions $X_0 = x$ and $\Gamma_0 = \gamma$. We let

$$T(x,\gamma,n) = \left\|A^{(n)}((x,\gamma),\cdot) - \pi(\cdot)\right\|$$

denote the total variation distance between the distribution of our adaptive algorithm at time n and the target distribution $\pi(\cdot)$. We call the adaptive algorithm *ergodic* if $\lim_{n\to\infty} T(x,\gamma,n) = 0$ for all $x \in \mathcal{X}$ and $\gamma \in \mathcal{Y}$.

Containment and Diminishing Adaptation ensure ergodicity and weak law of large number of adaptive MCMC, see [21].

Definition 2.1 (Containment). *for all $\epsilon > 0$, the sequence $\{M_\epsilon(X_n, \Gamma_n)\}_{n=0}^\infty$ is bounded in probability conditioned on $X_0 = x_*$ and $\Gamma_0 = \gamma_*$, where*

$$M_\epsilon(x,\gamma) = \inf_n \left\{n \geq 1 : \left\|P_\gamma^n(x,\cdot) - \pi(\cdot)\right\| \leq \epsilon\right\}$$

is the "ϵ-convergence function".

Definition 2.2 (Diminishing Adaptation)**.** $\lim_{n\to\infty} D_n = 0$ *in probability, where*

$$D_n = \sup_{x\in\mathcal{X}} \left\| P_{\Gamma_{n+1}}(x,\cdot) - P_{\Gamma_n}(x,\cdot) \right\|$$

is a $\mathcal{G}_{n+1}$ measurable random variable representing the amount of adapting done between iterations n and $n+1$.

Theorem 2.1 (RR [21])**.** *Consider an adaptive MCMC algorithm on a state space $\mathcal{X}$, with adaptation index $\mathcal{Y}$, so $\pi(\cdot)$ is stationary for each kernel P_γ for $\gamma \in \mathcal{Y}$. Assuming Containment and Diminishing Adaption, the adaptive algorithm is ergodic.*

Following standard results about geometric ergodicity and polynomial ergodicity, RR also considered certain "simultaneous" ergodicity conditions, as follows, see [22], [12], [6], [13], [9], [8], [1].

Definition 2.3 (simultaneously strongly aperiodically geometrically ergodic)**.** *Suppose that there is $C \in \mathcal{F}$, $V : \mathcal{X} \to [1,\infty)$, $\delta > 0, \lambda < 1$, and $b < \infty$, such that $\sup_C V = v < \infty$, and*
(i) $\exists$ a probability measure $\nu(\cdot)$ on C with $\mathbf{P}(x,\cdot) \geq \delta\nu(\cdot)$ for all $x \in C$; and
(ii) $PV \leq \lambda V + b\mathbf{1}_C$.

Theorem 2.2 (RR [21])**.** *Consider an adaptive MCMC algorithm with Diminishing Adaptation, such that the family $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ is simultaneously strongly aperiodically geometrically ergodic. Then the adaptive algorithm is ergodic.*

Results involving geometric convergence are well established, see [15], [14], [22], [12], [9]. The main method in these papers is to utilise Foster-Liapounov drift condition. From Theorem 2.1, we have the following:

Proposition 2.3. *Consider $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ a family of Markov chains on $\mathcal{X}$. Suppose that all compact sets are small for P_γ, $\gamma \in \mathcal{Y}$ and there exists a function V with $V > 1$ and $\sup_{x\in C,\gamma\in\mathcal{Y}} P_\gamma V(x) < \infty$ for all compact sets C:*

$$\limsup_{|x|\to\infty} \sup_{\gamma\in\mathcal{Y}} \frac{P_\gamma V(x)}{V(x)} < 1. \tag{2}$$

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

Proof: From Equation (2), 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 $V > 1$, $P_\gamma V(x) < \lambda V(x)$ for $|x| > K$. Since $\sup_{x\in C,\gamma\in\mathcal{Y}} P_\gamma V(x) < \infty$ for all compact sets, for any $\gamma \in \mathcal{Y}$, there exists a positive constant $b > 0$ such that $P_\gamma V(x) \leq \frac{\lambda+1}{2} V(x) + b\mathbf{1}_C$ for $\gamma \in \mathcal{Y}$. $\square$

Convergence with sub-geometric rates is studied using a sequence of drift conditions in [24]. It was shown by Jarner and Roberts in [13] 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

$$\mathbf{P}V \leq V - cV^\alpha + b\mathbf{1}_C, \tag{3}$$

then Markov chain converges to stationary distribution with a polynomial rate.

Proposition 2.4. *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 integer $m \in \mathbb{N}^+$, some constant $\delta > 0$, and some probability measure $\nu_\gamma(\cdot)$ on $\mathcal{X}$ such that $P_\gamma^m(x, \cdot) \geq \delta \mathbf{1}_C(x)\nu_\gamma(\cdot)$ for $\gamma \in \mathcal{Y}$. Suppose that there are some constants $\alpha \in (0, 1)$, $\beta \in (0, 1]$, $b > 0$, $c > 0$, and some measurable function $V(x) : \mathcal{X} \to [1, \infty)$ with $cV(x) > b$ on C^c, $\sup_{x \in C} V(x) < \infty$ and $\pi(V^\beta) < \infty$ such that*

$$P_\gamma V \leq V - cV^\alpha + b\mathbf{1}_C, \qquad \forall \gamma \in \mathcal{Y}. \tag{4}$$

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

Proof: see [4]. □

3 Counter Example

In this section, we introduce an example which shows that Diminishing Adaptation alone does not ensure ergodicity of adaptive MCMC. In fact, the example has $|\mathcal{Y}| = 2$, i.e. there are only finitely many different kernels P_γ.

Example 3.1. *Consider the Metropolis-Hastings algorithm with state space $\mathcal{X} = (0, \infty)$ and adaptive parameter space $\mathcal{Y} = \{-1, 1\}$, with the target density $\pi(x) \propto \frac{\mathbf{1}(x \geq 0)}{1+x^2}$. Let $\{Z_n\}$ be i.i.d. standard normal. The proposal values are given by $Y_n^{\Gamma_{n-1}} = X_{n-1}^{\Gamma_{n-1}} + Z_n$, 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 adaption is defined by $\Gamma_n = -\Gamma_{n-1}\mathbf{1}(X_n^{\Gamma_{n-1}} < \frac{1}{n}) + \Gamma_{n-1}\mathbf{1}(X_n^{\Gamma_{n-1}} \geq \frac{1}{n})$, 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 3.2. *The adaptive algorithm of Example 3.1 is not ergodic, i.e. X_n does not converge to the target distribution π.*

Proof: Assume that X_n converges to π. Define the hitting times:

$$\sigma_1 = \inf_n \{\Gamma_n \neq \Gamma_{n-1}\} \text{ and } \sigma_k = \inf_n \{n > \sigma_{k-1} : \Gamma_n \neq \Gamma_{n-1}\};$$
$$\tau_1 = \inf_n \{n > \sigma_1 : X_n \in (1/c, c)\} \wedge \sigma_2 \text{ and } \tau_k = \inf_n \{n > \sigma_k : X_n \in (1/c, c)\} \wedge \sigma_{k+1}.$$

Clearly, $\sigma_k < \tau_k \leq \sigma_{k+1}$.

First, we want to show that for any k, $P[\sigma_k < \infty] = 1$ implied by $E[\sum_{i=1}^\infty \mathbf{1}(\Gamma_i \neq \Gamma_{i-1})] = \infty$. Assume that there is some $m > 0$ such that $P[\sigma_m < \infty] \leq \epsilon < 1$. So, $P[\sum_{i=1}^\infty \mathbf{1}(\Gamma_i \neq \Gamma_{i-1}) \geq m] \leq \epsilon$. By induction we have that

$$\begin{aligned}
& P\left[\sum_{i=1}^\infty \mathbf{1}(\Gamma_i \neq \Gamma_{i-1}) \geq m(k+1)\right] \\
= \; & \int_{[\Gamma_n \neq \Gamma_{n-1}] \cap [\sum_{i=1}^n \mathbf{1}(\Gamma_i \neq \Gamma_{i-1}) = mk]} P[X_n \in dy]\, P_y\left[\sum_{i=1}^\infty \mathbf{1}(\Gamma_i \neq \Gamma_{i-1}) \geq m\right] \\
\leq \; & \epsilon P\left[\sum_{i=1}^\infty \mathbf{1}(\Gamma_i \neq \Gamma_{i-1}) \geq mk\right] \\
\leq \; & \epsilon^{k+1}.
\end{aligned}$$

Hence,

$$\begin{aligned}
E\left[\sum_{i=1}^{\infty}\mathbf{1}(\Gamma_i \neq \Gamma_{i-1})\right] &= \sum_{n=1}^{\infty} P\left[\sum_{i=1}^{\infty}\mathbf{1}(\Gamma_i \neq \Gamma_{i-1}) \geq n\right] \\
&\leq m\sum_{k=1}^{\infty} P\left[\sum_{i=1}^{\infty}\mathbf{1}(\Gamma_i \neq \Gamma_{i-1}) > km\right] \\
&\leq \frac{m}{1-\epsilon}.
\end{aligned}$$

On the other hand, by assumption, as n is large enough, $P\left[X_n \in A\right] \approx \pi(A)$ for any $A \in \mathcal{F}$ with $\pi(A) > 0$. So,

$$\begin{aligned}
P\left[\Gamma_n \neq \Gamma_{n-1}\right] &= \int P\left[\Gamma_n \neq \Gamma_{n-1} \mid X_{n-1} = x\right] P\left[X_{n-1} \in dx\right] \\
&\approx \int_0^{+\infty}\int_{-1/x}^{1/n-1/x} \varphi(z)dz \frac{2}{3.14(1+x^2)}dx \\
&\approx \frac{1}{n}\int_0^{\infty} \varphi(-1/x)\frac{2}{3.14(1+x^2)}dx = O\left(\frac{1}{n}\right).
\end{aligned}$$

Hence, $E\left[\sum_{i=1}^{\infty}\mathbf{1}(\Gamma_i \neq \Gamma_{i-1})\right] = \infty$. Therefore, for any k, $P\left[\sigma_k < \infty\right] = 1$.

Second, we consider the ratio of hitting the interval $(1/c,\, c)$ between σ_k and σ_{k+1}:

$$\begin{aligned}
\frac{E\left[\sum_{n=\sigma_k}^{\sigma_{k+1}}\mathbf{1}(X_n \in (1/c,\, c)) \mid \sigma_k\right]}{E\left[\sigma_{k+1}-\sigma_k \mid \sigma_k\right]} &\leq \frac{E\left[\sigma_{k+1}-\tau_k \mid \sigma_k\right]}{E\left[\sigma_{k+1}-\sigma_k \mid \sigma_k\right]} \\
&= \frac{E\left[\sigma_{k+1}-\sigma_k \mid \sigma_k\right] - E\left[\tau_k-\sigma_k \mid \sigma_k\right]}{E\left[\sigma_{k+1}-\sigma_k \mid \sigma_k\right]} \\
&\leq \frac{E\left[\sigma_{k+1}-\sigma_k \mid \sigma_k\right] - E\left[\tau_k-\sigma_k \mid \sigma_k\right]}{E\left[\tau_k-\sigma_k \mid \sigma_k\right]}.
\end{aligned}$$

Because the acceptance rate $\alpha(x,y) > 1$ if $0 < y < x$ under the adaptation $\gamma = 1$; $\alpha(x,y) > 1$ if $x < y$ under the adaptation $\gamma = -1$, the average time between σ_k and σ_{k+1} is less than the average time that one random-walk process hits to zero from above. So,

$$E\left[\sigma_{k+1}-\sigma_k \mid \sigma_k\right] \leq \sigma_k^2.$$

Taking enough large c, we can view the process X_n as random walk during the period $n \in (\sigma_k, \tau_k)$, because $\alpha(x,y) \approx 1$ when c is quite large. So,

$$E\left[\tau_k-\sigma_k \mid \sigma_k\right] \approx \left(\sigma_k - c\right)^2.$$

Hence,

$$\frac{E\left[\sum_{n=\sigma_k}^{\sigma_{k+1}}\mathbf{1}(X_n \in (1/c,\, c)) \mid \sigma_k\right]}{E\left[\sigma_{k+1}-\sigma_k \mid \sigma_k\right]} \leq \frac{\sigma_k^2 - (\sigma_k - c)^2}{(\sigma_k-c)^2} = \frac{2c\sigma_k - c^2}{(\sigma_k-c)^2}.$$

Finally, as k goes to ∞, σ_k and $\sigma_{k+1}-\sigma_k$ go to ∞, because $n - 1/n$ is increasing. So, the ratio of hitting $(1/c,\, c)$ between σ_k and σ_{k+1} should approximate to $\pi(1/c, c)$. However, $\frac{2c\sigma_k - c^2}{(\sigma_k-c)^2}$ tends to zero. Contradiction. □

Remark 3.1. From the proof, Diminishing Adaptation is satisfied. The algorithm is not ergodic so that Containment is not satisfied.

Remark 3.2. If the probability $P\left[\sigma_k = \infty\right] > 0$ for some k, the time that adaptation stays in one state is equal to infinity which leads the ergodicity of the algorithm, see the next section.

4 Summable Adaptive Condition

From the previous section, we know that Diminishing Adaptation is not sufficient for ergodicity. It was proved by Yang [25] that Adaptive MCMC is ergodic (and WLLN) assuming the conditions of Simultaneous Uniform Ergodicity and Summable Adaptive condition. Here, we will prove that a single Summable Diminishing Adaptation implies ergodicity of adaptive MCMC (without assuming Simultaneous Uniform Ergodicity). We also will present a modification of Example 3.1 which is ergodic.

Lemma 4.1. *Assume that $\mathcal{Y}$ is finite, and each P_γ is ergodic for $\pi(\cdot)$, and $\sum_{n=1}^{\infty} P(\Gamma_n \neq \Gamma_{n-1}) < \infty$. Then the adaptive algorithm is ergodic (i.e., converges to π).*

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

$$P\left(\Gamma_n = \Gamma_{n+1} = \cdots\right) > 1 - \epsilon/2. \tag{5}$$

Let $\mu_\epsilon := P_{\gamma_0} P_{\Gamma_1} \cdots P_{\Gamma_{N_0}}(x_0, \cdot)$. Since $\mathcal{Y}$ is finite, $\exists N_1(\epsilon) = N_1$, such that $\forall \gamma \in \mathcal{Y}$, $\forall n > N_1$,

$$\left\| \mu_\epsilon P_\gamma^n(x_0, \cdot) - \pi(\cdot) \right\| < \epsilon/2. \tag{6}$$

Taking $N = N_0 + N_1$, $\forall n > N$, we have that

$$\begin{aligned} \|\mathcal{L}(X_n) - \pi\| \leq \;& \sup_{A \in \mathcal{F}} \left| E\left[\mu_\epsilon P_{\Gamma_{N_0+1}} \cdots P_{\Gamma_{n-1}}(x_0, A) - \mu_\epsilon P_{\Gamma_{N_0}}^{n-N_0}(x_0, A) \mid X_0 = x_0, \Gamma_0 = \gamma_0 \right] \right| + \\ & \sup_{A \in \mathcal{F}} \left| E\left[\mu_\epsilon P_{\Gamma_{N_0}}^{n-N_0}(x_0, A) - \pi(A) \mid X_0 = x_0, \Gamma_0 = \gamma_0 \right] \right|. \end{aligned}$$

Since $n - N_0 > N_1$ and Equation (5),

$$\left| E\left[\mu_\epsilon P_{\Gamma_{N_0+1}} \cdots P_{\Gamma_{n-1}}(x_0, A) - \mu_\epsilon P_{\Gamma_{N_0}}^{n-N_0}(x_0, A) \mid X_0 = x_0, \Gamma_0 = \gamma_0 \right] \right| < \epsilon/2 * 2 = \epsilon.$$

By Equation (6),

$$\left| E\left[\mu_\epsilon P_{\Gamma_{N_0}}^{n-N_0}(x_0, A) - \pi(A) \mid X_0 = x_0, \Gamma_0 = \gamma_0 \right] \right| < \epsilon/2.$$

Therefore, $\|\mathcal{L}(X_n) - \pi\| < 3\epsilon/2$. □

Remark 4.1. The ergodicity assumption in Lemma 4.1 is not assumed to be uniformly bounded over choice of $\gamma \in \mathcal{Y}$.

Example 4.2. *Consider again the Metropolis-Hastings algorithm of Example 3.1, with $\mathcal{X} = (0, \infty)$ and $\mathcal{Y} = \{-1, 1\}$, and $\pi(x) \propto \frac{\mathbf{1}(x \geq 0)}{1+x^2}$, and is $Y_n^{\Gamma_{n-1}} = X_{n-1}^{\Gamma_{n-1}} + Z_n$ where $\{Z_n\}$ are i.i.d. standard normal. Assume now that the adaptive parameters $\{\Gamma_n\}$ are updated according to $\Gamma_n = -\Gamma_{n-1}\mathbf{1}(X_n^{\Gamma_{n-1}} < \frac{1}{n^{1+r}}) + \Gamma_{n-1}\mathbf{1}(X_n^{\Gamma_{n-1}} \geq \frac{1}{n^{1+r}})$ for some $r \geq 0$, so the case $r = 0$ corresponds to Example 3.1 (which was shown to be non-ergodic), while the case $r > 0$ is new.*

Proposition 4.3. *If $r > 0$, then the adaptive algorithm of Example 4.2 is ergodic, i.e. X_n converges to π.*

Proof: From the calculation in Example 3.1, we have that $P(\Gamma_n \neq \Gamma_{n-1} \mid X_{n-1} = x) = \int_{-\frac{1}{x}}^{\frac{1}{n^{1+r}} - \frac{1}{x}} \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) dz = O(\frac{1}{n^{1+r}})$. Therefore, $\sum_{n=1}^{\infty} P\left[\Gamma_n \neq \Gamma_{n-1}\right] < \infty$. So with σ_k in Example 3.1, we have $P\left[\sigma_k = \infty\right] > 0$ for some k. Hence, from Lemma 4.1, the adaptive algorithm is ergodic to π. □

Corollary 4.4. *Assume that each P_γ is ergodic for $\pi(\cdot)$, and that $\mathcal{Y}$ is compact in some topology with respect to which the mappings $(\gamma_1, \cdots, \gamma_k) \mapsto \|P_{\gamma_1} P_{\gamma_2} \cdots P_{\gamma_k}(x, \cdot) - \pi(\cdot)\|$ are all continuous for any fixed $x \in \mathcal{X}$ and $\gamma_1, \ldots, \gamma_k \in \mathcal{Y}$. If $\sum_{n=1}^{\infty} P(\Gamma_n \neq \Gamma_{n-1}) < \infty$ then adaptive algorithm is ergodic to π.*

Proof: We again follow the proof of Lemma 4.1. The only place where we used that $\mathcal{Y}$ was finite was to find N_1 such that Equation (6) holds for all $n > N_1$. But under the conditions of this corollary, we can use compactness to again find such N_1. □

5 Adaptive Metropolis Algorithms

The target density $\pi(\cdot)$ is defined on the state space $\mathcal{X} \subseteq \mathbb{R}^d$.

Assumption 5.1 (Target's 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.*

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, $n(z) := z/|z|$, and ∇ for the usual differential (gradient) operator.

Assumption 5.2 (Target's Strong Decrease)**.** *The target density π has continuous first derivatives and satisfies*

$$\limsup_{|x| \mapsto \infty} \langle n(x), m(x) \rangle < 0, \tag{7}$$

where $m(x) := \nabla \pi(x)/|\nabla \pi(x)|$.

Say adaptive MCMC is *adaptive Metropolis-Hastings algorithm* if for each $\gamma \in \mathcal{Y}$,

$$P_\gamma(x, dy) = \alpha(x, y) Q_\gamma(x, dy) + [1 - \alpha(x, y)]\, \delta_x(dy) \tag{8}$$

represents a Hastings algorithm with proposal measure $Q_\gamma(x, dy) = q_\gamma(x, y)\mu_d(dy)$, where $\alpha(x, y) := \frac{\pi(y) q_\gamma(y,x)}{\pi(x) q_\gamma(x,y)}$, and μ_d is Lebesgue measure.

Hastings algorithms are aperiodic and every compact set C with $\mu_d(C) > 0$ is small if target densities and the proposal densities are positive and continuous at very point, see [14]. This result was extend by Roberts and Tweedie in [22] that the Hastings Chain with proposal density $q_\gamma(x, y)$ is μ_d-irreducible and aperiodic, and every nonempty compact is small if the proposal density q_γ is locally positive.

Assumption 5.3 (Proposal's Local Positivity)**.** *There exist $\delta_\gamma > 0$ and $\epsilon_\gamma > 0$ such that*

$$q_\gamma(x) \geq \epsilon_\gamma, \text{ for } |x| \leq \delta_\gamma, \text{ for } \gamma \in \mathcal{Y}. \tag{9}$$

Assumption 5.4 (Proposal's Symmetry)**.** *Each proposal density in the proposal family has the form*

$$q_\gamma(x, y) = q_\gamma(x - y) = q_\gamma(y - x), \text{ for } \gamma \in \mathcal{Y}. \tag{10}$$

Say adaptive Metropolis-Hastings algorithm is *adaptive Metropolis algorithm* under Assumption 5.4. For each x in $\mathcal{X}$, define the *acceptance region* to be

$$A(x) = \{y \in \mathcal{X} | \pi(y) \geq \pi(x)\}, \tag{11}$$

and the *potential rejection region* to be

$$R(x) = \{y \in \mathcal{X} | \pi(y) < \pi(x)\}. \tag{12}$$

5.1 Target densities with light tails

For non-adaptive random-walk Metropolis algorithms, much is known about the connection between the tail behavior of the target density, and ergodicity properties of the algorithm. On $\mathbb{R}$, geometric convergence occurs essentially if and only if target density has geometric tails, see [14]. Some curvature conditions can be utilized to prove geometric ergodicity for a general class of target densities on $\mathbb{R}^d$ with tails at least as light as multivariate Gaussian, see [22]. Geometric ergodicity is equivalent to the acceptance probability being uniformly bounded away from zero, and if the target density is lighter-than-exponentially tailed and satisfies Assumption 5.2, then any random-walk-based Metropolis algorithm is geometrically ergodic, see [12]. In this section, we shall next consider how the tail property of target density affects Containment for adaptive Metropolis algorithms. We begin by considering target densities lighter-than-exponentially tailed. This class includes all multi-variate normal distributions. We begin with a definition.

Definition 5.1 (Lighter-than-exponential tail)**.** *The density $f(\cdot)$ on $\mathbb{R}^d$ is lighter-than-exponentially tailed if it is positive and has continuous first derivatives such that*

$$\limsup_{|x|\mapsto\infty} \langle n(x), \nabla \log f(x)\rangle = -\infty. \tag{13}$$

Remark 5.1. The definition implies that for any $r > 0$, there exists $R > 0$ such that

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

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

Remark 5.2. The normed gradient $m(x)$ will point towards the origin, while the direction $n(x)$ points away from the origin. For Definition 5.1, $\langle n(x), \nabla \log \pi(x)\rangle = \frac{|\nabla\pi(x)|}{\pi(x)} \langle n(x), m(x)\rangle$. Even if Assumption 5.2 holds, Equation (13) might not be true. E.g. $\pi(x) \propto \frac{1}{1+x^2}$, $x \in \mathbb{R}$. $m(x) = -n(x)\,|\nabla\pi|$ so that $\langle n(x), m(x)\rangle = -1$. $\langle n(x), \nabla \log \pi(x)\rangle = -\frac{2|x|}{1+x^2}$ so $\lim_{|x|\mapsto\infty} \langle n(x), \nabla \log \pi(x)\rangle = 0$.

Proposition 5.1. *If the target density π on $\mathbb{R}^d$ is normal (i.e. $N(\mu, \Sigma)$, Σ is positive definite), then π is strongly decreasing and lighter-than-exponentially tailed.*

Proof: Without loss of generalization, assume that $\mu = 0$.
Since $\pi(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{|\Sigma|^{1/2}} \exp(-x^\top \Sigma^{-1} x/2)$,

$$\langle n(x), m(x)\rangle = \left\langle \frac{x}{|x|}, \frac{-\Sigma^{-1}x}{|\Sigma^{-1}x|} \right\rangle = -\frac{x^\top \Sigma^{-1} x}{|x|\,|\Sigma^{-1}x|}.$$

Since Σ is a real symmetric and positive definite matrix, suppose that $\Sigma = A^\top D A$ where A is orthogonal, and D is diagonal with positive diagonal elements. Hence,

$$\frac{x^\top \Sigma^{-1} x}{|x|\,|\Sigma^{-1}x|} = \frac{y D^{-1} y}{|y|\,|D^{-1}y|} = \frac{\sum_{i=1}^d y_i^2 d_i^{-1}}{\sqrt{\sum_{i=1}^d y_i^2 \sum_{i=1}^d d_i^{-2} y_i^2}} \geq \frac{\min\left(d_i^{-1}\right)}{\max\left(d_i^{-1}\right)}.$$

where $y = Ax$.

$$\langle n(x), \nabla \log \pi(x)\rangle = \frac{|\nabla\pi(x)|}{\pi(x)} \left\langle \frac{x}{|x|}, \frac{-\Sigma^{-1}x}{|\Sigma^{-1}x|} \right\rangle = -\frac{x D^{-1} x}{|x|} \underset{|x|\to\infty}{\to} -\infty.$$

So, the result holds. □

We now consider target densities with exponential tails. We shall show under some conditions that for target densities exponentially tailed on $\mathbb{R}^d$, adaptive Metropolis algorithm is ergodic.

Definition 5.2 (Exponential tail)**.** *The density function $f(\cdot)$ on $\mathbb{R}^d$ is exponentially tailed if it is a positive, continuously differentiable function on $\mathbb{R}^d$, and*

$$\limsup_{|x|\to\infty} \langle n(x), \nabla \log f(x)\rangle < 0. \tag{14}$$

Remark 5.3. There exists $\beta > 0$ such that for x sufficiently large,

$$\langle n(x), \nabla \log f(x)\rangle = \langle n(x), n(\nabla f(x))\rangle \, |\nabla \log f(x)| \leq -\beta.$$

Further, if $0 < -\langle n(x), n(\nabla f(x))\rangle \leq 1$, then $|\nabla \log f(x)| \geq \beta$.

Before giving our result, we state Lemma 4.2 in [12].

Lemma 5.2. *Let x and z be two distinct points in $\mathbb{R}^d$, and let $\xi = n(x-z)$. If*

$$\langle \xi, m(y)\rangle \neq 0$$

for all y on the line from x to z, then z does not belong to $\left\{y \in \mathbb{R}^d : \pi(y) = \pi(x)\right\}$.

Assumption 5.5. *Suppose the target density π is exponentially tailed and strongly decreasing (Assumption 5.2), and each proposal distribution $Q_\gamma(\cdot,\cdot)$ for $\gamma \in \mathcal{Y}$ is symmetric (Assumption 5.4). Define $\eta_1 := -\limsup_{|x|\to\infty} \langle n(x), m(x)\rangle$ and $\eta_2 := -\limsup_{|x|\to\infty} \langle n(x), \nabla \log \pi(x)\rangle$.*
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 for any sequence $\{(x_n, \gamma_n)\}$ with $|x_n| \to +\infty$ and $\{\gamma_n\} \subset \mathcal{Y}$, $\exists$ subsequence $\{(x_{n_k}, \gamma_{n_k})\}$ with $|x_{n_k}| \to \infty$ such that

$$\lim_{k\to\infty} \int_{\left\{z = a\xi \,|\, \delta \leq a \leq \Delta,\, \xi \in S^{d-1},\, \left|\xi - n(x_{n_k})\right| < \epsilon/3\right\}} |z| \, q_{\gamma_{n_k}}(z) \mu_d(dz) > \frac{3}{\beta\epsilon(e-1)}, \tag{15}$$

where S^{d-1} be the unit hypersphere in $\mathbb{R}^d$, and $a\xi$ represents the scalar multiple of the vector $\xi \in \mathbb{R}^d$ by $a \in \mathbb{R}$.

Remark 5.4. Since the integral in Equation (15) depends on the direction $n(x)$ of x, not on the length $|x|$. The criteria in the assumption is equivalent to

$$\inf_{(u,\gamma)\in S^{d-1}\times\mathcal{Y}} \int_{\left\{z = a\xi \,|\, \delta \leq a \leq \Delta,\, \xi \in S^{d-1},\, |\xi - u| < \epsilon/3\right\}} |z| \, q_\gamma(z) \mu_d(dz) > \frac{3}{\beta\epsilon(e-1)}. \tag{16}$$

Lemma 5.3. *Suppose that the target density π is exponentially tailed and smooth enough with $\eta_1 := -\limsup_{|x|\to\infty} \langle n(x), m(x)\rangle$ and $\eta_2 := -\limsup_{|x|\to\infty} \langle n(x), \nabla \log \pi(x)\rangle$; the proposal family $\{Q_\gamma(\cdot)\}_{\gamma\in\mathcal{Y}}$ is symmetric; there is a function $q^-(z) := g(|z|)$, $q^-(\cdot) : \mathcal{X} \to \mathbb{R}^+$ and $g(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$, such that there is $M \geq 0$, for $|z| \geq M$, $q_\gamma(z) \geq q^-(z)$ for $\gamma \in \mathcal{Y}$.*
If there are $\epsilon \in (0, \eta_1)$ and $\frac{1}{\eta_2} \vee M < \delta < \Delta$ such that

$$\frac{(d-1)\pi^{\frac{d-1}{2}}}{2\Gamma(\frac{d+1}{2})} B_{r^2}\left(\frac{d-1}{2}, \frac{1}{2}\right) \int_\delta^\Delta g(t) t^d dt > \frac{3}{\eta_1\eta_2(e-1)}, \tag{17}$$

where $r := \frac{\epsilon}{6}\sqrt{36-\epsilon^2}$, and the incomplete beta function $B_x(\alpha,\beta) := \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt$, then Assumption 5.5 holds.

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

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

where $\omega(\cdot)$ denotes the surface measure on S^{d-1}.
By the symmetry of $u \in S^{d-1}$, let $u = e_d$. So, the projection from the piece $\{\xi \in S^{d-1} : |\xi - u| < \epsilon/3\}$ of the hypersphere S^{d-1} to the subspace $\mathbb{R}^{d-1}$ generated by the first $d-1$ coordinates is $d-1$ hyperball $V^{d-1}(0,r)$ with the center 0 and the radius $r = \frac{\epsilon}{6}\sqrt{36-\epsilon^2}$. Define $f(z) = \sqrt{1-(z_1^2+\cdots+z_{d-1}^2)}$.

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

Hence,

$$\int_{\{z=a\xi \mid \delta \le a \le \Delta,\, \xi\in S^{d-1},\, |\xi-u|<\epsilon/3\}} |z|\, g(|z|)\mu_d(dz) = \frac{(d-1)\pi^{\frac{d-1}{2}}}{2\Gamma(\frac{d+1}{2})}B_{r^2}\left(\frac{d-1}{2},\frac{1}{2}\right)\int_\delta^\Delta g(t)t^d dt. \quad (18)$$

Therefore, the result holds. □

Consider the test function $V_s(x) = c\pi^{-s}(x)$ for some $c > 0$ and $s \in (0,1)$ such that $V(x) \ge 1$. By some algebras,

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

From Proposition 3 in RR [18], we have $P_\gamma V_s(x)/V_s(x) \le r(s)V_s(x)$ where $r(s) := 1+s(1-s)^{1/s-1}$.

Proposition 5.4 (Exponential tail)**.** *Suppose that the target density π is exponentially tailed, regular (Assumption 5.1), and strongly decreasing (Assumption 5.2). Consider an adaptive Metropolis algorithm (Assumption 5.4) with the proposal family $\{Q_\gamma(\cdot,\cdot)\}_{\gamma\in\mathcal{Y}}$ of which each proposal density is locally positive (Assumption 5.3). If Assumption 5.5 holds, then Containment holds.*

Proof: Consider the measurable function $V(x) := c\pi^{-s}(x)$ for $s \in (0,1)$. By Assumption 5.1, $V(x) \ge 1$ for some constant c, and for any compact set $C \subset \mathcal{X}$, $\sup_{x\in C} V(x) < \infty$ so that $\sup_{x\in C,\gamma\in\mathcal{Y}} P_\gamma V(x) < \infty$. Since $\frac{P_\gamma V(x)}{V(x)} = \frac{P_\gamma \pi^{-s}(x)}{\pi^{-s}(x)}$ and Proposition 2.3, it is sufficient to show that $\limsup_{|x|\to\infty}\sup_{\gamma\in\mathcal{Y}}\frac{P_\gamma\pi^{-s}(x)}{\pi^{-s}(x)} < 1$.
Assume that $\limsup_{|x|\to\infty}\sup_{\gamma\in\mathcal{Y}}\frac{P_\gamma\pi^{-s}(x)}{\pi^{-s}(x)} \ge 1$. So, there exists a sequence $\{(x_n,\gamma_n)\}$ with $|x_n| \to \infty$ and $\{\gamma_n\} \subset \mathcal{Y}$ such that $\lim_{n\to\infty}\frac{P_{\gamma_n}\pi^{-s}(x_n)}{\pi^{-s}(x_n)} \ge 1$.
From Assumption 5.5, there exist $\epsilon \in (0,\eta_1)$, $\beta \in (0,\eta_2)$ (η_1 and η_2 are defined in the assumption), δ, and Δ such that there exists a subsequence $\{x_{n_k}\}$ with $\lim_{k\to\infty}|x_{n_k}| = \infty$ such that the properties in the assumption are satisfied, and $\lim_{k\to\infty}\frac{P_{\gamma_{n_k}}\pi^{-s}(x_{n_k})}{\pi^{-s}(x_{n_k})} \ge 1$.

We denote the cones by

$$C(x_{n_k}) := C_{n_k} := \Big\{x_{n_k} - a\xi \mid \delta \le a \le \Delta,\, \xi \in S^{d-1},\, |\xi - n(x_{n_k})| \le \epsilon/3\Big\}. \tag{19}$$

Denote the set of points of C_{n_k} rotated 180° degrees about x_{n_k} by

$$C^r(x_{n_k}) : C^r_{n_k} := \Big\{x_{n_k} + a\xi \mid \delta \le a \le \Delta,\, \xi \in S^{d-1},\, |\xi - n(x_{n_k})| \le \epsilon/3\Big\}. \tag{20}$$

There exists $N_1 > 0$ such that $\forall k > N_1$, $|x_{n_k}| > 2\Delta$. So, for $y \in C_{n_k} \cup C^r_{n_k}$ (i.e. $y = x_{n_k} \pm a\xi$ for some $\xi \in S^{d-1}$ and some $a \in (\delta, \Delta)$), $|y| \ge |x_{n_k}| - \Delta > \Delta$ so that

$$|n(y) - n(x_{n_k})| < |\xi - n(x_{n_k})| \le \epsilon/3.$$

From Definition 5.2, there exists $K_1 > 0$ such that $|x| > K_1$, $\langle n(x), \nabla \log \pi(x)\rangle \le -\beta$. So, there exists $N_2 \in \mathbb{N}^+$, for $k > N_2$, $|x_{n_k}| > K_1$.
From Assumption 5.2, there exists $K_2 > K_1$ such that $|x| > K_2$, $\langle n(x), m(x)\rangle \le -\epsilon$. So, there exists $N_3 \in \mathbb{N}^+$, for $k > N_3$, $C_{n_k} \cup C^r_{n_k} \subset \{z \in \mathbb{R}^d : |z| > K\}$, so that $\langle n(y), m(y)\rangle \le -\epsilon$, for $y \in C_{n_k} \cup C^r_{n_k}$.
Then, for $k > N_1 \vee N_2 \vee N_3$ and $y \in C_{n_k} \cup C^r_{n_k}$,

$$\langle \xi, m(y)\rangle = \langle \xi - n(x_{n_k}), m(y)\rangle + \langle n(x_{n_k}) - n(y), m(y)\rangle + \langle n(y), m(y)\rangle < -\epsilon/3, \tag{21}$$

and

$$|\nabla \log \pi(y)| = \frac{\langle n(y), \nabla \log \pi(y)\rangle}{\langle n(y), m(y)\rangle} > \beta. \tag{22}$$

Hence, by Lemma 5.2,

$$C_{n_k} \cap \Big\{y \in \mathbb{R}^d : \pi(y) = \pi(x_{n_k})\Big\} = \emptyset \text{ and } C^r_{n_k} \cap \Big\{y \in \mathbb{R}^d : \pi(y) = \pi(x_{n_k})\Big\} = \emptyset.$$

For $y = x_{n_k} - a\xi \in C_{n_k}$,

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

So, $C_{n_k} \subset A(x_{n_k})$. By similar technique, $C^r_{n_k} \subset R(x_{n_k})$.
Consider the test function $V_s(x) = \pi^{-s}(x)$. We have

$$\begin{aligned}
P_{\gamma_{n_k}} V_s(x_{n_k})/V_s(x_{n_k}) &= \int_{\{C_{n_k} - x_{n_k}\}\cup\{C^r_{n_k} - x_{n_k}\}} I_{x_{n_k},s}(z) q_{\gamma_{n_k}}(z)\mu_d(dz) + \\
&\quad \int_{\{C_{n_k} - x_{n_k}\}^c\cap\{C^r_{n_k} - x_{n_k}\}^c} I_{x_{n_k},s}(z) q_{\gamma_{n_k}}(z)\mu_d(dz),
\end{aligned}$$

where

$$I_{x_{n_k},s}(z) = \begin{cases} \frac{\pi^s(x_{n_k})}{\pi^s(x_{n_k}+z)}, & z \in A(x_{n_k}) - x_{n_k}, \\ 1 - \frac{\pi(x_{n_k}+z)}{\pi(x_{n_k})} + \frac{\pi^{1-s}(x_{n_k}+z)}{\pi^{1-s}(x_{n_k})}, & z \in R(x_{n_k}) - x_{n_k}. \end{cases} \tag{23}$$

For $z = a\xi \in C^r_{n_k} - x_{n_k}$, by Definition 5.2, Equations (21) and (22),

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

So, by Assumption 5.5,

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

Similarly, for $z = -a\xi \in C_{n_k} - x_{n_k}$,

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

Since $t^{1-s} - t$ is an increasing function on $[0, 1/e]$ for $s \in (0, 1)$,

$$\begin{aligned}
& \int_{\left\{C_{n_k} - x_{n_k}\right\} \cup \left\{C^r_{n_k} - x_{n_k}\right\}} I_{x_{n_k},s}(z) q_{\gamma_{n_k}}(z)\mu_d(dz) \\
= & \int_{C_{n_k} - x_{n_k}} \frac{\pi^s(x_{n_k})}{\pi^s(x_{n_k} + z)} q_{\gamma_{n_k}}(z)\mu_d(dz) + \\
& \int_{C^r_{n_k} - x_{n_k}} \left(1 - \frac{\pi(x_{n_k} + z)}{\pi(x_{n_k})} + \frac{\pi^{1-s}(x_{n_k} + z)}{\pi^{1-s}(x_{n_k})}\right) q_{\gamma_{n_k}}(z)\mu_d(dz) \\
\le & \int_{C_{n_k} - x_{n_k}} e^{-s\beta\epsilon|z|/3} q_{\gamma_{n_k}}(z)\mu_d(dz) + \int_{C^r_{n_k} - x_{n_k}} \left(1 - e^{-\beta\epsilon|z|/3} + e^{-(1-s)\beta\epsilon|z|/3}\right) q_{\gamma_{n_k}}(z)\mu_d(dz).
\end{aligned}$$

Since $\sup_{t\in(0,1)} \left(1 - t + t^{1-s}\right) \le 1 + s(1-s)^{1/s-1} \le 1 + se^{-1+s}$, $0 \le I_{x_{n_k},s}(z) \le 1 + se^{-1+s}$.

$$\begin{aligned}
& \int_{\left\{C_{n_k} - x_{n_k}\right\}^c \cap \left\{C^r_{n_k} - x_{n_k}\right\}^c} I_{x_{n_k},s}(z) q_{\gamma_{n_k}}(z)\mu_d(dz) \\
\le & \left(1 + se^{-1+s}\right) Q_{\gamma_{n_k}}\left(\left\{C_{n_k} - x_{n_k}\right\}^c \cap \left\{C^r_{n_k} - x_{n_k}\right\}^c\right).
\end{aligned}$$

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

$$H_{x,\gamma}(\theta, s) := K_{x,\gamma}(s\theta) + K_{x,\gamma}(0) - K_{x,\gamma}(\theta) + K_{x,\gamma}((1-s)\theta) + \left(1 + se^{-1+s}\right)(1 - 2K_{x,\gamma}(0)). \quad (24)$$

So,

$$P_{\gamma_{n_k}} V_s(x_{n_k})/V_s(x_{n_k}) \le H_{x_{n_k},\gamma_{n_k}}(\beta\epsilon/3, s).$$

Thus, by simple algebra, we have that

$$\begin{aligned}
H_{x_{n_k},\gamma_{n_k}}(\beta\epsilon/3, 0) &= 1, \\
\frac{\partial H_{x_{n_k},\gamma_{n_k}}(\beta\epsilon/3, 0)}{\partial s} &= e^{-1}(1 - 2K_{x_{n_k},\gamma_{n_k}}(0)) - \frac{\beta\epsilon}{3}\int_{C_{n_k} - x_{n_k}} |z|\, q_{\gamma_{n_k}}(z)\mu_d(dz) + \\
& \quad \frac{\beta\epsilon}{3}\int_{C_{n_k} - x_{n_k}} |z|\, e^{-\beta\epsilon|z|/3} q_{\gamma_{n_k}}(z)\mu_d(dz) \\
&\le e^{-1} - \frac{\beta\epsilon}{3}\left(1 - e^{-1}\right)\int_{C_{n_k} - x_{n_k}} |z|\, q_{\gamma_{n_k}}(z)\mu_d(dz).
\end{aligned}$$

The cone $C_{n_k} - x_{n_k}$ is only dependent on the unit vector $n(x_{n_k})$ and the angle degree ϵ.
From Assumption 5.5,

$$\limsup_{k\to\infty} \frac{\partial H_{x_{n_k},\gamma_{n_k}}(\beta\epsilon/3, 0)}{\partial s} < 0.$$

So, there exists $s \in (0,1)$ such that $\lim_{k\to\infty} P_{\gamma_{n_k}} V_s(x_{n_k})/V_s(x_{n_k}) < 1$, which leads to contradiction. By Proposition 2.3, Containment holds. □

Theorem 5.5. *Under the conditions described either in Proposition 5.4, if Diminishing Adaptation holds then the adaptive Metropolis algorithm is ergodic.*

Proof: By Theorem 2.1, the result holds. □

When a target density is lighter-than-exponentially tailed, it is also exponentially tailed. Here we present one relatively relaxed assumption for target with lighter-than-exponentially tailed density. For Assumption 5.5, we need to find two finite positive value δ and Δ greater than $\frac{3}{\beta\epsilon}$. However, for density lighter-than-exponentially tailed, β can be arbitrary large positive value even infinity so δ can be taken as arbitrarily small positive value even zero.

Assumption 5.6. *Suppose the target density π is lighter-than-exponentially tailed and strongly decreasing (Assumption 5.2), and each proposal distribution $Q_\gamma(\cdot,\cdot)$ for $\gamma \in \mathcal{Y}$ is symmetric (Assumption 5.4). Define $\eta := -\limsup_{|x|\to\infty} \langle n(x), m(x)\rangle$.*
Assume that there are $\epsilon \in (0,\eta)$, $0 < \delta < \Delta \le \infty$ such that for any sequence $\{(x_n,\gamma_n)\}$ with $|x_n| \to +\infty$ and $\{\gamma_n\} \subset \mathcal{Y}$, $\exists$ subsequence $\{(x_{n_k},\gamma_{n_k})\}$ with $|x_{n_k}| \to \infty$ such that

$$\liminf_{k\to\infty} \int_{\left\{z=a\xi \,|\, \delta<a<\Delta,\, \xi\in S^{d-1},\, \left|\xi - n(x_{n_k})\right|<\epsilon/3\right\}} q_{\gamma_{n_k}}(z)\mu_d(dz) > 0, \tag{25}$$

where S^{d-1} be the unit hypersphere in $\mathbb{R}^d$, and $a\xi$ represents the scalar multiple of the vector $\xi \in \mathbb{R}^d$ by $a \in \mathbb{R}$.

Remark 5.5. Since the integral in Equation (25) depends on the direction $n(x)$ of x, not on the length $|x|$. The criteria in the assumption is equivalent to

$$\inf_{(u,\gamma)\in S^{d-1}\times\mathcal{Y}} \int_{\left\{z=a\xi \,|\, \delta<a<\Delta,\, \xi\in S^{d-1},\, |\xi-u|<\epsilon/3\right\}} q_\gamma(z)\mu_d(dz) > 0. \tag{26}$$

Lemma 5.6. *Suppose that the target density π is lighter-than-exponentially tailed, and strongly decreasing (Assumption 5.2). Consider an adaptive Metropolis algorithm (Assumption 5.4) with the proposal family $\{Q_\gamma(\cdot,\cdot)\}_{\gamma\in\mathcal{Y}}$. Suppose further that there exists $M > 0$ such that for $|z| > M$, there exists a positive function $q^-(\cdot)$ such that for any $\gamma \in \mathcal{Y}$, $q_\gamma(z)\mathbf{1}(|z| > M) \ge q^-(z)\mathbf{1}(|z| > M) > 0$. Then Assumption 5.6 holds.*

Proof: Let $\delta = M$.

$$\begin{aligned} &\int_{\left\{z=a\xi \,|\, \delta<a<\infty,\, \xi\in S^{d-1},\, |\xi-u|<\epsilon/3\right\}} q_\gamma(z)\mu_d(dz) \\ > \;& Q^-\left(\left\{a\xi \;|\; M < a < \infty, \xi \in S^{d-1}. \, |\xi - u| < \epsilon/3\right\}\right) > 0, \end{aligned}$$

So, Assumption 5.6 holds. □

Theorem 5.7 (Lighter-than-exponential tail)**.** *Suppose that the target density π is lighter-than-exponentially tailed, regular (Assumption 5.1), and strongly decreasing (Assumption 5.2). Consider an adaptive Metropolis algorithm (Assumption 5.4) with the proposal family $\{Q_\gamma(\cdot,\cdot)\}_{\gamma\in\mathcal{Y}}$ of which each proposal density is locally positive (Assumption 5.3). If Assumption 5.6 and Diminishing Adaptation holds, then the algorithm is ergodic.*

Proof: Consider the measurable function $V(x) := c\pi^{-s}(x)$ for $s \in (0,1)$. By Assumption 5.1, $V(x) \geq 1$ for some constant c, and for any compact set $C \subset \mathcal{X}$, $\sup_{x\in C} V(x) < \infty$ so that $\sup_{x\in C,\gamma\in\mathcal{Y}} P_\gamma V(x) < \infty$. Since $\frac{P_\gamma V(x)}{V(x)} = \frac{P_\gamma \pi^{-s}(x)}{\pi^{-s}(x)}$ and Proposition 2.3, it is sufficient to show that $\limsup_{|x|\to\infty} \sup_{\gamma\in\mathcal{Y}} \frac{P_\gamma\pi^{-s}(x)}{\pi^{-s}(x)} < 1$.
Assume that $\limsup_{|x|\to\infty} \sup_{\gamma\in\mathcal{Y}} \frac{P_\gamma\pi^{-s}(x)}{\pi^{-s}(x)} \geq 1$. So, there exists a sequence $\{(x_n,\gamma_n)\}$ with $|x_n| \to \infty$ and $\{\gamma_n\} \subset \mathcal{Y}$ such that $\lim_{n\to\infty} \frac{P_{\gamma_n}\pi^{-s}(x_n)}{\pi^{-s}(x_n)} \geq 1$.
From Assumption 5.6, there exists $\epsilon \in (0,\eta)$, $0 < \delta < \Delta \leq \infty$ such that there exists a subsequence $\{(x_{n_k},\gamma_{n_k})\}$ with $\lim\limits_{k\to\infty} |x_{n_k}| = \infty$ such that the properties in the assumption are satisfied, and $\lim\limits_{k\to\infty} \frac{P_{\gamma_{n_k}}\pi^{-s}(x_{n_k})}{\pi^{-s}(x_{n_k})} \geq 1$.
Since π is lighter-than-exponentially tailed, for any $\beta > 0$, there exists $K > 0$ such that $|x| > K$, $\langle n(x), \nabla \log \pi(x)\rangle < -\beta$. Using the similar technique in the proof of Proposition 5.4, for sufficiently large $k(\beta)$, $\delta \geq \frac{3}{\epsilon\sqrt{\beta}}$,

$$\frac{\pi(x_{n_k}+z)}{\pi(x_{n_k})} \leq e^{-\beta\epsilon|z|/3} \leq e^{-\sqrt{\beta}} \text{ for } z \in C^r_{n_k} - x_{n_k}, \text{ and}$$

$$\frac{\pi(x_{n_k})}{\pi(x_{n_k}+z)} \leq e^{-\beta\epsilon|z|/3} \leq e^{-\sqrt{\beta}} \text{ for } z \in C_{n_k} - x_{n_k},$$

where C_{n_k} and $C^r_{n_k}$ are defined in Equations (19) and (20), and depend on δ and Δ.
Consider the sequence β_j with $\lim\limits_{j\to+\infty} \beta_j = +\infty$. So, there is a subsequence $\{(x_{n_{k_j}},\gamma_{n_{k_j}})\}$ such that the above equations hold. However,

$$\int_{C_{n_{k_j}}-x_{n_{k_j}}} e^{-s\beta_j\epsilon|z|/3} q_{\gamma_{n_{k_j}}}(z)\mu_d(dz) \leq e^{-s\sqrt{\beta_j}},$$

$$\int_{C^r_{n_{k_j}}-x_{n_{k_j}}} e^{-(1-s)\beta_j\epsilon|z|/3} q_{\gamma_{n_{k_j}}}(z)\mu_d(dz) \leq e^{-(1-s)\sqrt{\beta_j}}.$$

Hence,

$$\lim_{j\to\infty} K_{x_{n_{k_j}},\gamma_{n_{k_j}}}(s\beta_j\epsilon/3) = 0, \text{ and } \lim_{j\to\infty} K_{x_{n_{k_j}},\gamma_{n_{k_j}}}((1-s)\beta_j\epsilon/3) = 0,$$

where $K_{x,\gamma}(\cdot)$ is defined in the proof of Proposition 5.4.
So,

$$\lim_{j\to\infty} P_{\gamma_{n_{k_j}}} V_s(x_{n_{k_j}})/V_s(x_{n_k}) < 1 + se^{-1+s} - (1+2se^{-1+s}) \liminf_{j\to\infty} K_{x_{n_{k_j}},\gamma_{n_{k_j}}}(0).$$

From Assumption 5.6, for some $s \in (0,1)$, $\lim_{j\to\infty} P_{\gamma_{n_{k_j}}} V_s(x_{n_{k_j}})/V_s(x_{n_k}) < 1$. Therefore, by Proposition 2.3, Containment holds. By Theorem 2.1, the result holds. □

Here we discuss two examples. The first one (Example 5.8) is from RR [20] 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.* [11]. In the example, the target density has lighter-than-exponential tails. The second (Example 5.11) concerns with target densities with truly exponential tails.

Example 5.8. *Consider a d-dimensional target distribution $\pi(\cdot)$ which is regular, strongly decreasing and lighter-than-exponentially tailed. 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 $n > 2d$,*

$$Q_n(x,\cdot) = \begin{cases} (1-\theta)N(x,(2.38)^2\Sigma_n/d) + \theta N(x,(0.1)^2 I_d/d), & \Sigma_n \text{ is positive definite,} \\ N(x,(0.1)^2 I_d/d), & \Sigma_n \text{ is not positive definite,} \end{cases} \tag{27}$$

for some fixed $\theta \in (0,1)$, and the empirical covariance matrix

$$\Sigma_n = \frac{1}{n}\left(\sum_{i=0}^{n} X_i X_i^\top - (n+1)\overline{X}_n \overline{X}_n^\top\right), \tag{28}$$

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 5.6. The proposal $N(x,(2.38)^2\Sigma/d)$ is optimal in a particular large-dimensional context, see [17] and [19]. Thus the proposal $N(x,(2.38)^2\Sigma_n/d)$ is an effort to approximate this.

Remark 5.7. Commonly, the iterative form of Equation (28) is more useful,

$$\Sigma_n = \frac{n-1}{n}\Sigma_{n-1} + \frac{1}{n+1}\left(X_n - \bar{X}_{n-1}\right)\left(X_n - \bar{X}_{n-1}\right)^\top. \tag{29}$$

Proposition 5.9. *Consider an adaptive Metropolis algorithm (Assumption 5.4) with the proposal family $\{Q_\gamma(\cdot,\cdot)\}_{\gamma\in\mathcal{Y}}$ of which each proposal density is locally positive (Assumption 5.3).*

Suppose that the target density π is exponentially tailed, regular (Assumption 5.1), and strongly decreasing (Assumption 5.2).

If Assumption 5.5 is satisfied and the algorithm's adaptive scheme is defined as that in Example 5.8, then Diminishing Adaptation holds.

Proof: From Proposition 5.4, the family $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ is simultaneously strongly aperiodically geometrically ergodic with the test function $V(x) = c\pi^{-s}(x)$ for some $s \in (0,1)$ and some $c > 0$. So, 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,

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

Hence,

$$\begin{aligned} & \|\Sigma_n - \Sigma_{n-1}\|_M \\ \le \; & \tfrac{1}{n+1}\left\|X_n X_n^\top\right\|_M + \tfrac{1}{n-1}\left\|\tfrac{1}{n}\textstyle\sum_{i=0}^{n-1} X_i X_i^\top\right\|_M + \tfrac{2}{n}\left\|\overline{X}_{n-1}\overline{X}_{n-1}^\top\right\|_M + \\ & \tfrac{1}{n+1}\left\|X_n \overline{X}_{n-1}^\top + \overline{X}_{n-1} X_n^\top\right\|_M. \end{aligned} \tag{30}$$

To prove $\Sigma_n - \Sigma_{n-1}$ converges to zero in probability, it is sufficient to check that $\left\|X_n X_n^\top\right\|_M$, $\left\|\frac{1}{n}\sum_{i=0}^{n-1} X_i X_i^\top\right\|_M$, $\left\|\overline{X}_{n-1}\overline{X}_{n-1}^\top\right\|_M$ 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 5.1. If $r \geq 2$ then $\frac{r-1}{r} \geq 0.5$. Therefore, as $|x|$ is extremely large, $V(x) \geq |x|^2$.
Since $\left\|X_n X_n^\top\right\|_M := \sup_{|u|=1} u^\top X_n X_n^\top u \leq \sup_{|u|=1} |u|^2 |X_n|^2 \leq |X_n|^2$, $\left\|X_n X_n^\top\right\|_M$ is bounded in probability.
Obviously,

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

Then, for $K > 0$,

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

We know that $\sup_n E[V(X_n)] < \infty$ (See Theorem 18 in [21]). Hence, $\left\|\frac{1}{n}\sum_{i=0}^{n-1} X_i X_i^\top\right\|_M$ is bounded in probability.
$\left|\overline{X}_n\right| \leq \frac{1}{n+1}\sum_{i=0}^{n} |X_i|$. So,

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

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

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

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

Theorem 5.10. *The algorithm of Example 5.8 is ergodic.*

Proof: Obviously, the proposal densities has uniformly lower bound function. From the proof of Lemma 5.6, Assumption 5.6 holds. By Theorem 5.7 and Proposition 5.9, the adaptive Metropolis algorithm is ergodic. □

Example 5.11. *Consider the standard multivariate double 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*

$$Q_n(x,\cdot) = \begin{cases} Unif\left(V^d(x,\Delta)\right), & n \le 2d, \text{ or } \Sigma_n \text{ is nonsingular}, \\ (1-\theta)N(x,(2.38)^2\Sigma_n/d) + \theta\, Unif\left(V^d(x,\Delta)\right), & n > 2d, \text{ and } \Sigma_n \text{ is singular}, \end{cases} \tag{31}$$

for $\theta \in (0,1)$, $Unif\left(V^d(x,\Delta)\right)$ *is an uniform distribution on the hyperball* $V^d(x,\Delta)$ *with the center* x *and the radius* Δ, *and* Σ_n *is as defined in Equation (28). The problem is: how to choose* Δ *such that the adaptive Metropolis algorithm is ergodic?*

Proposition 5.12. *There exists a large enough* $\Delta > 0$ *such that the adaptive Metropolis algorithm of Example 5.11 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 regular, exponentially tailed, and strongly decreasing. Obviously, each proposal density is locally positive. Now, let us check Assumption 5.5 by using Lemma 5.3. Because

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

the function $g(t)$ defined in Lemma 5.3 is equal to $\frac{1}{Vol(V^d(x,\Delta))}$. The parameters η_1 and η_2 defined in Lemma 5.3 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 Equation (17) is

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

where $B(x,y)$ and $B_r(x,y)$ are beta function and incomplete beta function, r is a function of ϵ defined in Lemma 5.3.

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 Equation (17) holds. By Lemma 5.3, Assumption 5.5 holds. Then, by Proposition 5.4, Containment holds. By Proposition 5.9, Diminishing Adaptation holds. By Theorem 2.1, the adaptive Metropolis algorithm is ergodic. □

5.2 Target densities with heavy tails

Now, we consider a particular class of target densities with tails which are heavier than exponential. It was previously shown by Fort and Moulines [7] 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 5.3 (Hyperbolic tail)**.** *The density function* $f(\cdot)$ *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* $|x|$,

$0 < d_0|x|^m \le -\log f(x) \le D_0|x|^m$;
$0 < d_1|x|^{m-1} \le |\nabla \log f(x)| \le D_1|x|^{m-1}$;
$0 < d_2|x|^{m-2} \le \left|\nabla^2 \log f(x)\right| \le D_2|x|^{m-2}$.

Assumption 5.7 (Proposal's Uniform Compact Support)**.** *There exists a* $M > 0$ *such that for any* $\gamma \in \mathcal{Y}$ *and* $|z| > M$, $q_\gamma(z) = 0$.

Say that the proposal family has *Uniform Upper Bound density* if there is a positive function $q^+(\cdot)$ with $\int q^+(z)\mu_d(dz) < \infty$, such that for any $\gamma \in \mathcal{Y}$, $q_\gamma(\cdot) \le q^+(\cdot)$. Denote $Q^+(dz) = q^+(z)\mu_d(dz)$.

Theorem 5.13. *Suppose that the target density* π *is hyperbolically tailed, regular (Assumption 5.1), and strongly decreasing (Assumption 5.2). Consider an adaptive Metropolis algorithm (Assumption 5.4) with the proposal family* $\{Q_\gamma(\cdot,\cdot)\}_{\gamma\in\mathcal{Y}}$ *of which each proposal density is locally positive (Assumption 5.3), and has Uniform Upper Bound function and Uniform Compact Support (Assumption 5.7). If Diminishing Adaptation holds, the adaptive algorithm is ergodic.*

Proof: From Assumptions 5.1, 5.4, and 5.7, each P_γ is ergodic to π. By the definition of Uniform Upper Bounded density, we can first find $D > 0$ such that $\int_{|x|>D} q^+(x)dx < 1$ and then let $q^*(x) = q^+(x)$ for $|x| > D$, and define $q^*(x)$ for $|x| < D$ as necessary to make q^* have integral 1. Then q^* is a density, and $q_\gamma(x) \le q^*(x)$ for all γ and all $|x| > D$. Denote $Q^*(dz) = q^*(z)\mu_d(dz)$. Since q_γ has uniform bounded support, we can assume q^* with bounded support.
We consider the test function $V_s(x) = (-\log \pi(x))^s$:

$$\begin{aligned}
P_\gamma V_s(x) - V_s(x) &= \int \left[V_s(x+z)\alpha(x,x+z) + V_s(x)(1-\alpha(x,x+z))\right] q_\gamma(z)\mu_d(dy) - V_s(x) \\
&\le \int_{|z|<D} \left[V_s(x+z)\alpha(x,x+z) + V_s(x)(1-\alpha(x,x+z))\right] (q_\gamma - q^*)(z)\mu_d(dy) + \\
&\quad P^*V_s(x) - V_s(x) \\
&\le Q^*(|z|\ge D)V_s(x) + \\
&\quad \int_{|z|<D} \left[V_s(x+z) - V_s(x)\right] \alpha(x,x+z)(q_\gamma - q^*)(z)\mu_d(dz) + \\
&\quad P^*V_s(x) - V_s(x).
\end{aligned}$$

On the other hand, $\nabla V_s(x) = sV_{s-1}\nabla V_1(x)$. Letting

$$R(s,x,z) = V_s(x+z) - V_s(x) - sV_{s-1}(x)\langle \nabla V_1(x), z\rangle,$$

$$\begin{aligned}
\sup_{|z|<M\wedge\delta_\gamma} |R(s,x,z)|\,|z|^{-2} &\le \sup_{|z|<M\wedge\delta_\gamma} \left|\nabla^2 V_s(z)\right| \\
&\le s \sup_{|z|<M\wedge\delta_\gamma} V_{s-2}(z) \left|(s-1)\nabla V_1(z)\nabla V_1(z)^\top + V_1(z)\nabla^2 V_1(z)\right|.
\end{aligned}$$

Since the target density is hyperbolic tailed,

$$\limsup_{|x|\to\infty} |x|^{2-sm} \sup_{|z|<M\wedge\delta_\gamma} |R(s,x,z)|\,|z|^{-2} < \infty, \tag{32}$$

So, by Assumption 5.7,

$$\int_{|z|<D} |R(s,x,z)|\, q_\gamma(z)\mu_d(dz) \le \int_{|z|<D\wedge M} |R(s,x,z)|\, q^+(z)\mu_d(dz) = O\left(|x|^{ms-2}\right).$$

By the symmetry of $q_\gamma(\cdot)$ and $q^*(\cdot)$,

$$\int_{|z|<D} sV_{s-1}(x) \langle \nabla V_1(x), z \rangle (q_\gamma - q^*)(z)\mu_d(dz) = 0.$$

Hence,

$$\begin{aligned} &\int_{|z|<D} [V_s(x+z) - V_s(x)]\, \alpha(x, x+z) (q_\gamma - q^*)(z)\mu_d(dz) \\ \le &\int_{|z|<D\wedge M} |R(s,x,z)| \left(q^+ + q^*\right)(z)\mu_d(dz) = O\left(|x|^{ms-2}\right). \end{aligned}$$

Thus, taking $s < (2-m)/m$,

$$\limsup_{|x|\to\infty} \sup_{\gamma\in\mathcal{Y}} \lim_{D\to\infty} \int_{|z|<D} [V_s(x+z) - V_s(x)]\, \alpha(x, x+z) (q_\gamma - q^*)(z)\mu_d(dz) = 0.$$

Since $\delta_\gamma \le M$ where δ_γ is defined in Assumption 5.7, letting that D goes to infinity,

$$\lim_{D\to\infty} Q^*(|z| \ge D)V_s(x) = 0.$$

Therefore,

$$P_\gamma V_s(x) - V_s(x) \le P^* V_s(x) - V_s(x).$$

Since the proposal distribution Q^* satisfies the property in Assumption 5.7, P^* is ergodic, converging to π. By Proposition 2.4, Containment holds. Hence, by Theorem 2.1, the adaptive Metropolis algorithm is ergodic. □

6 Adaptive Metropolis-within-Gibbs Algorithms

We now consider so-called *Metropolis-within-Gibbs* algorithms, which update each of the d coordinates separately according to its own Metropolis algorithm.

For $\gamma \in \mathcal{Y}$, let $(P_{1,\gamma}, \dots, P_{d,\gamma})$ be any collection of Markov kernels on the state space $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_d \subseteq \mathbb{R}^d$. The adaptive random scan hybrid sampler for the collection is the sampler $P_{\mathrm{RS},\gamma}$ defined by

$$P_{\mathrm{RS},\gamma} := d^{-1}\left(P_{1,\gamma} + \cdots + P_{d,\gamma}\right),$$

where each $P_{i,\gamma}$ arises from a symmetric random-walk Metropolis algorithm on the ith coordinate with the proposal distribution $Q_{i,\gamma}(x, dy) = q_{i,\gamma}(x,y)\mu(dy)$. We require various assumptions.

Assumption 6.1 (Local Positivity). *There exist $\delta_{i,\gamma} > 0$ and $\epsilon_{i,\gamma} > 0$ such that*

$$q_{i,\gamma}(z) \ge \epsilon_{i,\gamma}, \text{ for } |z| \le \delta_{i,\gamma}. \tag{33}$$

Assumption 6.2 (Symmetry). *For each coordinate i, $q_{i,\gamma}(x, x+ze_i) = q_{i,\gamma}(x, x-ze_i) := q_{i,\gamma}(z)$ for $\gamma \in \mathcal{Y}$ where e_i is the unit vector of the coordinate i.*

The transition kernels $P_{i,\gamma}, i \in \{1,\dots,d\}$, on $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ are defined as follows: for $x = (x_1,\dots,x_d) \in \mathbb{R}^d$, $A = A_1 \times \cdots \times A_d \in \mathcal{B}\left(\mathbb{R}^d\right)$, $z \in \mathbb{R}$,

$$\begin{aligned} P_{i,\gamma}(x, A) := &\prod_{k\ne i} \delta_{x_k}(A_k) \int_{A_i - x_i} \alpha(x, x+ze_i) q_{i,\gamma}(z)\mu(dz) \\ &+\delta_x(A) \int \left(1 - \alpha(x, x+ze_i)\right) q_{i,\gamma}(z)\mu(dz), \end{aligned}$$

where $A_i - x_i := \{z \in \mathbb{R}, x_i + z \in A_i\}$ and $\alpha(x, x + ze_i) := 1 \wedge \frac{\pi(x+ze_i)}{\pi(x)}$. Let $A(x,i)$ and $R(x,i)$ be the acceptance region and potential rejection region respectively in the ith direction:

$$A(x,i) = \{z \in \mathbb{R} : \pi(x + ze_i) \geq \pi(x)\},$$
$$R(x,i) = \{z \in \mathbb{R} : \pi(x + ze_i) < \pi(x)\}.$$

For adaptive Metropolis-within-Gibbs algorithms, we mainly adapt the method of Fort *et al.* [9]. First, we restate Proposition 2 in Fort *et al.* [9]:

Proposition 6.1. *Under Assumptions 5.4 and 5.1, let $V_s(x) = \pi^{-s}(x)$ for $s \in (0,1)$. For all $x \in \mathbb{R}^d$,*

$$P_{i,\gamma}V_s(x) \leq r(s)V_s(x), \tag{34}$$

where

$$r(s) := 1 + s(1-s)^{1/s-1}.$$

It can also be shown that

$$\frac{P_{i,\gamma}V_s(x)}{V_s(x)} = \int I(z,x,i,s)q_{i,\gamma}(z)\mu(dz),$$

where

$$I(z,x,i,s) := \begin{cases} \left(\frac{\pi(\tilde{x}^j)}{\pi(\tilde{x}^j + z\,e_i)}\right)^s, & z \in A(x,i), \\ 1 - \frac{\pi(\tilde{x}^j + z\,e_i)}{\pi(\tilde{x}^j)} + \left(\frac{\pi(\tilde{x}^j + z\,e_i)}{\pi(\tilde{x}^j)}\right)^{1-s}, & z \in R(x,i). \end{cases} \tag{35}$$

Assumption 6.3. *There is an $\beta > 0$ and a δ such that $1/\beta \leq \delta < \Delta \leq \infty$, for any sequence (x^j, γ^j) with $\lim_j = |x^j| = +\infty$ and $\{\gamma^j\} \subset \mathcal{Y}$, we may extract a subsequence $(\tilde{x}^j, \tilde{\gamma}^j)$ with the property that, for some $i \in \{1, \ldots, d\}$, we have for all $z \in [\delta, \Delta]$,*

$$\lim_j \frac{\pi(\tilde{x}^j)}{\pi(\tilde{x}^j - sign(\tilde{x}^j_i)ze_i)} \leq \exp(-\beta z) \text{ and } \lim_j \frac{\pi(\tilde{x}^j + sign(\tilde{x}^j_i)ze_i)}{\pi(\tilde{x}^j)} \leq \exp(-\beta z); \tag{36}$$

Moreover,

$$\liminf_{j\to\infty} \inf_{i\in\{1,\ldots,d\}} \int_\delta^\Delta zq_{i,\tilde{\gamma}^j}(z)\mu(dz) > \frac{d}{\beta(e-1)}. \tag{37}$$

Remark 6.1. Equation (36) means that on each coordinate, the tail of the target density π decays exponentially. Equation (37) represents the relationship of the first moment of proposal density and the decaying rate of the target density tails. The property in Equation (37) is equivalent to

$$\inf_{\gamma\in\mathcal{Y}} \inf_{i\in\{1,\ldots,d\}} \int_\delta^\Delta zq_{i,\gamma}(z)\mu(dz) > \frac{d}{\beta(e-1)}.$$

These two equations are not difficult to check, see Examples 6.4 and 6.6. Furthermore, there are some similarities between this assumption and Assumption 5.5.

Adapting the procedure of Theorem 3 in Fort *et al.* [9], we have the following result for adaptive Metropolis-within-Gibbs algorithms applied to exponentially tailed target distributions.

Theorem 6.2 (Exponential tail)**.** *Suppose that the target density π is regular (Assumption 5.1). Consider a random-walk based Metropolis-within-Gibbs algorithm with the family of proposal distributions $\{Q_{i,\gamma}(\cdot,\cdot)\}_{i=1,\cdots,d;\gamma\in\mathcal{Y}}$ where all proposal densities are symmetric (Assumption 6.2) and locally positive (Assumption 6.1). If Assumption 6.3 holds, the adaptive algorithm with Diminishing Adaptation is ergodic.*

Proof: Assume that for any $s\in(0,1)$, $\limsup_{|x|\to\infty}\sup_{\gamma\in\mathcal{Y}} P_{\mathrm{RS},\gamma}V_s(x)/V_s(x)\ge 1$. Then, there exists a sequence pair $\{(x^j,\gamma^j)\}$ with $\lim_{j\to\infty}\left|x^j\right|\to\infty$ and $\{\gamma^j\}\subset\in\mathcal{Y}$ such that $\lim_{j\to\infty}P_{\mathrm{RS},\gamma_j}V_s(x_j)/V_s(x_j)\ge 1$.

Under Assumption 6.3, we may extract from the sequence (x^j,γ^j) a subsequence $(\tilde{x}^j,\tilde{\gamma}^j)$ such that for some $i\in\{1,\dots,d\}$, Equations (36) and (37) are satisfied. Without loss of generality, assume $sign(\tilde{x}^j_i)=1$. Let $J(\delta,\Delta)=[-\Delta,-\delta]\cup[\delta,\Delta]$. It is easy to prove that

$$\lim_j R(\tilde{x}^j,i)\cap J(\delta,\Delta)=[\delta,\Delta] \text{ and } \lim_j A(\tilde{x}^j,i)\cap J(\delta,\Delta)=[-\Delta,-\delta].$$

So, since $u^{1-s}-u$ is an increasing function on $[0,1/e]$ for $s\in(0,1)$, by Assumption 6.3,

$$\begin{aligned}
&\int_{J(\delta,\Delta)} I(z,\tilde{x}^j,i,s)q_{i,\tilde{\gamma}^j}(z)\mu(dz)\\
=\ &\int_{\{A(\tilde{x}^j,i)\cap J(\delta,\Delta)\}}\left(\frac{\pi\left(\tilde{x}^j\right)}{\pi\left(\tilde{x}^j+z\,e_i\right)}\right)^s q_{i,\tilde{\gamma}^j}(z)\mu(dz)+\\
&\int_{\{R(\tilde{x}^j,i)\cap J(\delta,\Delta)\}}\left[1+\left(\frac{\pi\left(\tilde{x}^j+z\,e_i\right)}{\pi\left(\tilde{x}^j\right)}\right)^{1-s}-\frac{\pi\left(\tilde{x}^j+z\,e_i\right)}{\pi\left(\tilde{x}^j\right)}\right]q_{i,\tilde{\gamma}^j}(z)\mu(dz)\\
\le\ &K_{i,\tilde{\gamma}^j}(\beta s)+K_{i,\tilde{\gamma}^j}(0)+K_{i,\tilde{\gamma}^j}(\beta(1-s))-K_{i,\tilde{\gamma}^j}(\beta)
\end{aligned}$$

where $K_{i,\gamma}(t)=\int_{\delta\vee 1/\beta}^{\Delta}e^{-tz}q_{i,\gamma}(z)\mu(dz)$. Hence,

$$\begin{aligned}
&P_{\mathrm{RS},\tilde{\gamma}^j}V_s(\tilde{x}^j)/V_s(\tilde{x}^j)\\
\le\ &\frac{1}{d}(K_{i,\tilde{\gamma}^j}(\beta s)+K_{i,\tilde{\gamma}^j}(0)+K_{i,\tilde{\gamma}^j}(\beta(1-s))-K_{i,\tilde{\gamma}^j}(\beta)+r(s)(1-2K_{i,\tilde{\gamma}^j}(0)))+\frac{d-1}{d}r(s)\\
\le\ &\frac{1+se^{-1+s}}{d}(d-2K_{i,\tilde{\gamma}^j}(0))+\frac{1}{d}(K_{i,\tilde{\gamma}^j}(\beta s)+K_{i,\tilde{\gamma}^j}(0)+K_{i,\tilde{\gamma}^j}(\beta(1-s))-K_{i,\tilde{\gamma}^j}(\beta))\\
=\ &H_{i,\tilde{\gamma}^j}(\beta,s)
\end{aligned}$$

where $H_{i,\gamma}(\beta,s)=\frac{1+se^{-1+s}}{d}(d-2K_{i,\gamma}(0))+\frac{1}{d}(K_{i,\gamma}(\beta s)+K_{i,\gamma}(0)+K_{i,\gamma}(\beta(1-s))-K_{i,\gamma}(\beta))$. By simple algebra, we know the following things:

$$\begin{aligned}
H_{i,\tilde{\gamma}^j}(\beta,0)&=1;\\
\frac{\partial H_{i,\tilde{\gamma}^j}}{\partial s}(\beta,0)&=(d-2K_{i,\tilde{\gamma}^j}(0))e^{-1}-\beta\int_\delta^\Delta zq_{i,\tilde{\gamma}^j}(z)\mu(dz)+\beta\int_\delta^\Delta ze^{-\beta z}q_{i,\tilde{\gamma}^j}(z)\mu(dz)\\
&\le d/e-\beta(1-1/e)\int_\delta^\Delta zq_{i,\tilde{\gamma}^j}(z)\mu(dz).
\end{aligned}$$

By Assumption 6.3,

$$\limsup_{j\to\infty}\frac{\partial H_{i,\tilde{\gamma}^j}}{\partial s}(\beta,0)<0.$$

Therefore, $\limsup_j H_{i,\tilde{\gamma}^j}(\beta,s)<1$ for some $s\in(0,1)$, which leads to a contradiction. □

Continuing Assumption 6.3, whenever the coordinate tails of a target density decay lighter-than-exponentially, i.e. β in Assumption 6.3 is equal to infinity, the δ can be arbitrarily small positive value even zero, and the Δ can be arbitrarily large positive value even infinity. Moreover, the left hand side of Equation (37) is obviously great than zero, because each proposal has density and the density is symmetric. Hence, we have the following assumption and result.

Assumption 6.4. *There exist $0 \le \delta < \Delta \le \infty$ such that for any sequence (x^j, γ^j) with $lim_{j\to\infty}|x^j| = +\infty$ and $\{\gamma^j\} \in \mathcal{Y}$, we may extract a subsequence $(\tilde{x}^j, \tilde{\gamma}^j)$ with the property that, for some $i \in \{1, \ldots, d\}$ and all $z \in (0, \infty)$,*

$$\lim_{j\to\infty} \frac{\pi(\tilde{x}^j)}{\pi(\tilde{x}^j - sign(\tilde{x}^j_i)ze_i)} = 0 \text{ and } \lim_{j\to\infty} \frac{\pi(\tilde{x}^j + sign(\tilde{x}^j_i)ze_i)}{\pi(\tilde{x}^j)} = 0. \tag{38}$$

Moreover,

$$\liminf_{j\to\infty} \inf_{i\in\{1,\ldots,d\}} \int_\delta^\Delta q_{i,\tilde{\gamma}^j}(z)\mu(dz) > 0. \tag{39}$$

Theorem 6.3 (Lighter-than-exponential tails)**.** *Suppose that the target density π is regular (Assumption 5.1). Consider a random-walk based Metropolis-within-Gibbs algorithm with the family of proposal distributions $\{Q_{i,\gamma}(\cdot,\cdot)\}_{i=1,\cdots,d;\gamma\in\mathcal{Y}}$ where all proposal densities are symmetric (Assumption 6.2) and locally positive (Assumption 6.1). If Assumption 6.4 holds, the adaptive algorithm with Diminishing Adaptation is ergodic.*

Proof: Assume that for any $s \in (0,1)$, $\limsup_{|x|\to\infty} \sup_{\gamma\in\mathcal{Y}} P_{\mathrm{RS},\gamma}V_s(x)/V_s(x) \ge 1$. Then, there exists a sequence pair $\{(x^j, \gamma^j)\}$ with $\lim_{j\to\infty} \left|x^j\right| \to \infty$ and $\{\gamma^j\} \subset\in \mathcal{Y}$ such that $\lim_{j\to\infty} P_{\mathrm{RS},\gamma_j}V_s(x_j)/V_s(x_j) \ge 1$.

Consider $\beta_j \uparrow +\infty$. From Assumption 6.4, we may extract from the sequence (x^j, γ^j) a subsequence $(\tilde{x}^j, \tilde{\gamma}^j)$ such that for all $z \in (0, \infty)$,

$$\frac{\pi(\tilde{x}^j)}{\pi(\tilde{x}^j - sign(\tilde{x}^j_i)ze_i)} < e^{-\sqrt{\beta_j}} \text{ and } \lim_{j\to\infty} \frac{\pi(\tilde{x}^j + sign(\tilde{x}^j_i)ze_i)}{\pi(\tilde{x}^j)} < e^{-\sqrt{\beta_j}}.$$

Hence,

$$\lim_{j\to\infty} K_{i,\tilde{\gamma}^j}(\beta_j s) = 0, \text{ and } \lim_{j\to\infty} K_{i,\tilde{\gamma}^j}(\beta_j(1-s)) = 0,$$

where $K_{i,\gamma}(t)$ is defined in the proof of Theorem 6.2.
So,

$$\lim_{j\to\infty} P_{\mathrm{RS},\tilde{\gamma}^j}V_s(\tilde{x}^j)/V_s(\tilde{x}^j) \le 1 + se^{-1+s} - \frac{1+2se^{-1+s}}{d} \liminf_{j\to\infty} K_{i,\tilde{\gamma}^j}(0).$$

As s goes to zero, se^{-1+s} goes to zero. By Equation (39), for $s \in (0,1)$, $\lim_{j\to\infty} P_{\mathrm{RS},\tilde{\gamma}^j}V_s(\tilde{x}^j)/V_s(\tilde{x}^j) < 1$. By Proposition 2.3, Containment holds. By Theorem 2.1, the result holds. □

Example 6.4. *We consider the mixed Gaussian density on $\mathbb{R}^2$. Define*

$$\pi(x) = \beta\exp(-(x_1^2 + x_2^2)) + (1-\beta)exp(-(x_1^2 + x_1^2x_2^2 + x_2^2)),$$

where $\beta \in [0,1]$. Consider random-walk based adaptive Metropolis-within-Gibbs algorithm with proposal family $Q_{i,\gamma}(\cdot,\cdot)$ where on each coordinate, the proposal density is normal, and their minimal variance is $\epsilon > 0$. Assume that Diminishing Adaptation is satisfied.

Proposition 6.5. *The algorithm of Example 6.4 is ergodic.*

Proof: We have that

$$\frac{\nabla_1 \log \pi(x)}{-2x_1} = \frac{\beta \exp(-(x_1^2+x_2^2)) + (1+x_2^2)(1-\beta)\exp(-(x_1^2+x_1^2x_2^2+x_2^2))}{\pi(x)} \in \left[1, 1+x_2^2\right],$$
$$\frac{\nabla_2 \log \pi(x)}{-2x_2} = \frac{\beta \exp(-(x_1^2+x_2^2)) + (1+x_1^2)(1-\beta)\exp(-(x_1^2+x_1^2x_2^2+x_2^2))}{\pi(x)} \in \left[1, 1+x_1^2\right].$$

Clearly, $\nabla_i \log \pi(x)/(-2x_i)$ is positive bounded. So, Equation (38) is satisfied. The target density is lighter-than-exponentially tailed on each coordinate.

$$\int_\epsilon^\infty \frac{z}{\sqrt{2\pi\sigma^2}} \exp(-z^2/2\sigma^2)dz = \sigma \int_{\epsilon/\sigma}^\infty \frac{z}{\sqrt{2\pi}} \exp(-z^2/2)dz \geq \epsilon \int_1^\infty \frac{z}{\sqrt{2\pi}} \exp(-z^2/2)dz.$$

Since, all density has minimal variance ϵ, Equation (39) holds. Thus, by Theorem 6.3, the algorithm is ergodic. □

Finally, we consider the target density of Example 8 in [9], a mixture of two exponential distributions.

Example 6.6. *For some $a > 1$, define*

$$\pi(x) \propto 0.5e^{-|x_1|-a|x_2|} + 0.5e^{-a|x_1|-|x_2|}, \qquad x = (x_1, x_2).$$

Consider random-walk based adaptive Metropolis-within-Gibbs algorithm with proposal family $Q_{i,\gamma}(\cdot,\cdot)$ where on each coordinate i, the proposal distribution is a mixed distribution of Uniform$(-b,b)$ and $N(0,\sigma_i^2)$ respectively with weights ϵ and $1-\epsilon$. Assume that Diminishing Adaptation is satisfied. Then, for sufficient large $b > 0$, the adaptive algorithm is ergodic.

Proposition 6.7. *For sufficiently large $b > 0$, the algorithm of Example 6.6 is ergodic.*

Proof: It is sufficient to check Containment. The β defined in Assumption 6.3 is a. It is easy to check that Equations (36) holds.

$$\epsilon \int_\delta^b \frac{z}{2b} dz = \frac{\epsilon(b^2-\delta^2)}{4b}.$$

Let $\delta = 1 \in (0, a)$. So, we only need to take suitable b such that

$$\frac{\epsilon(b^2-\delta^2)}{4b} > \frac{2}{a(e-1)}.$$

Then Equation (37) holds so Assumption 6.3 holds. □

7 Conclusions and Discussions

For adaptive Metropolis and adaptive Metropolis-within-Gibbs algorithms, we provide some conditions only related to properties of the target density and the proposal family. Under these conditions, the adaptive algorithms converge in the fast sense (simultaneously strongly aperiodically geometrically ergodic). Generally speaking, target densities is required to be regular, strongly decreasing, and at least exponentially tailed. However, for adaptive Metropolis-within-Gibbs algorithms, the target density is only required to be exponentially tailed on the direction of coordinates, and strong decrease is not needed. For truly exponentially tailed target densities, we found that ergodicity of adaptive algorithms is related to the dimensions of the state space. Especially, for adaptive Metropolis algorithms, if proposal densities have uniform lower bound function, then ergodicity of algorithms is connected to the d^{th} moment of the function on some hypercone on $\mathbb{R}^d$.

Recently, there also is some results about this topic, see Saksman and Vihola [23]. They show that if the target density is regular, strongly decreasing, and strongly lighter-than-exponentially tailed ($\limsup_{|x|\to\infty} \langle n(x), \nabla \log \pi(x)\rangle / |x|^{\rho-1} = -\infty$ for some $\rho > 1$) which is used to keep the convexity of outside manifold contour of target densities, then SLLN for symmetric random-walk based adaptive Metropolis algorithms holds. Compared with our results, although our conditions do not require that the target density is strongly lighter-than-exponentially tailed, one restriction on proposal density is needed. On the other hand, RR [21] give one example (Example 24) that shows that for some adaptive chain, SLLN does not hold, but ergodicity of the chain holds.

Jarner and Hansen [12] show that if target density is lighter-than-exponential tailed and strongly decreasing then random-walk-based Metropolis algorithm is geometrically ergodic. The technique in Proposition 5.4 can be also applied to MCMC. So, even if target density is exponentially tailed under some moment condition similar as Equation (15), 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 different from the assumption ($q(x, y) = q(|x - y|)$) of Jarner and Hansen [12]. Our assumption generalizes theirs.

Acknowledgements

We thank M. Vihola for helpful comments.

References

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

[2] Andrieu, C and Robert, C.P. Controlled mcmc for optimal sampling. *Preprint*, 2002.

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

[4] Bai, Y. Simultaneous drift conditions on adaptive Markov Chain Monte Carlo algorithms. *Preprint*, 2009.

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

[6] Fort, G. and Moulines, E. Computable bounds for subgeometrical and geometrical ergodicity. 2000.

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

[8] Fort, G. and Moulines, E. Polynomial ergodicity of Markov transition kernels. *Stoch. Process. Appl.*, 103:57–99, 2003.

[9] Fort, G., and Moulines, E., Roberts, G.O. and Rosenthal, J.S. On the geometric ergodicity of hybrid samplers. *J. Appl. Prob.*, 40:123–146, 2003.

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

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

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

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

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

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

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

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

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

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

[20] Roberts, G.O. and Rosenthal, J.S. Examples of Adaptive MCMC. *Preprint*, 2006.

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

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

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

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

[25] Yang, C. On the weak law of large number for unbounded functionals for adaptive MCMC. *Preprint*, 2008.