ON THE STABILITY AND ERGODICITY OF AN ADAPTIVE SCALING METROPOLIS ALGORITHM

MATTI VIHOLA

Abstract. This paper considers the stability and ergodicity of an adaptive random walk Metropolis algorithm. The algorithm adjusts the scale of the symmetric proposal distribution continuously, based on the observed acceptance probability. A strong law of large numbers is shown to hold for functionals bounded on compact sets and growing at most exponentially as $\|x\| \to \infty$, assuming that the target density is smooth enough and has either compact support or super-exponentially decaying tails.

1. Introduction

Markov chain Monte Carlo (MCMC) is a general method often used to approximate integrals of the type

$$I := \int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x < \infty$$

where π is a probability density function [see, e.g., 9, 14, 17]. The method is based on a Markov chain $(X_n)_{n\geq 1}$ that can be simulated in practice, and for which $I_n := \sum_{k=1}^{n} f(X_k) \to I$ as $n \to \infty$. Such a chain can be constructed, for example, as follows. Assume q is a zero-mean Gaussian probability density, and let $X_1 \equiv x_1$ for some fixed point $x_1 \in \mathbb{R}^d$. For $n \geq 2$, recursively,

(S1) simulate $Y_n = X_{n-1} + \theta W_n$, where W_n are independent random vectors distributed according to q, and

(S2) with probability $\alpha_n = \min\{1, \pi(Y_n)/\pi(X_{n-1})\}$ the proposal is accepted and $X_n = Y_n$; otherwise the proposal is rejected and $X_n = X_{n-1}$.

This symmetric random-walk Metropolis algorithm will produce, with any positive scalar parameter θ, a valid chain, i.e. that $I_n \to I$ almost surely as $n \to \infty$ [see, e.g. 13, Theorem 1]. However, the efficiency of the method, i.e. the speed of the convergence $I_n \to I$, is crucially affected by the choice of θ. For too large θ, very few proposals become accepted, and the chain mixes poorly. For too small θ, most

---

Date: March 24, 2009.

2000 *Mathematics Subject Classification.* Primary 65C40; Secondary 60J27, 93E15, 93E35.

Key words and phrases. Adaptive Markov chain Monte Carlo, ergodicity, Metropolis algorithm, stability, stochastic approximation.

The author was supported by the Academy of Finland, projects no. 110599 and 201392, by the Finnish Academy of Science and Letters, Vilho, Yrjö, and Kalle Väisälä Foundation, by the Finnish Centre of Excellence in Analysis and Dynamics Research, and by the Finnish Graduate School in Stochastics and Statistics.

of the proposals Y_n become accepted, but the steps $X_n - X_{n-1}$ are small, preventing good mixing. In fact, previous results indicate that the acceptance probability is closely related with the efficiency of the algorithm. The "rule of thumb" is that the acceptance probability α_n should be on the average about 0.234 [18, 19], although this choice is not always optimal [8]. In practice, such a θ is usually found by several trial runs, which can be laborious and time-consuming.

So called adaptive MCMC algorithms have gained popularity since the seminal work of Haario, Saksman, and Tamminen [10]. Several other such algorithms have been proposed after Andrieu and Robert [2] noticed the connection between Robbins-Monro stochastic approximation and adaptive MCMC [1, 3, 6, 15, 16]. The Adaptive Scaling Metropolis (ASM) algorithm considered in this paper optimises the scaling of the proposal distribution adaptively, based on the observed acceptance probability [3, 5, 6, 16]. Namely, in the step (S1) of the above algorithm, θ is replaced with a random variable θ_{n-1} set initially to $\theta_1 > 0$, and for $n \geq 2$ defined through recursion

(S3) $\log \theta_n = \log \theta_{n-1} + c n^{-\gamma}(\alpha_n - \alpha^*)$

where α^* is the desired mean acceptance probability, e.g. $\alpha^* = 0.234$, and $c > 0$ and $\gamma \in (1/2, 1]$ are constants.

It is not obvious that such an algorithm is valid, i.e. that $I_n \to I$. In fact, there are examples of adaptive MCMC schemes that destroy the correct ergodic properties [15]. To ensure the validity of such an algorithm, it is essential that the effect of the adaptation "diminishes" in some sense as $n \to \infty$. For many practical algorithms, however, showing the adaptation to diminish turns out to be a difficult task. Current ergodicity results on adaptive MCMC algorithms assume some "uniform" behaviour for all the possible MCMC kernels [5, 6, 15]. In the context of the ASM algorithm, this means essentially that θ_n must be restricted to a predefined set $[a, b]$ with some $0 < a \leq b < \infty$. Alternatively, one can use a general "reprojection and reinitialisation" technique with a sequence of such sets $[a_n, b_n]$ with $a_n \searrow 0$ and $b_n \nearrow \infty$ as proposed by Andrieu and Moulines [1], or stabilisation methods that modify the adaptation rule to ensure stable behaviour [3]. Roberts and Rosenthal [16] have also suggested using a fixed "non-adaptive" proposal distribution component within the adaptive scheme; [7] provides analysis on this alternative.

It is a common belief that many of the proposed adaptive MCMC algorithms, including the ASM algorithm described above, are inherently stable and thereby do not require such artificial restrictions or stabilisation structures. There is some empirical evidence of the stability, but few theoretical results. In particular, Saksman and Vihola [20] verify that the Adaptive Metropolis algorithm has the correct ergodic properties and is stable, provided the target distribution π has super-exponentially decaying tails with regular contours. The next section shows that the stability and ergodicity of the ASM algorithm can be verified under similar conditions fairly easily, without any artificial stabilisation.

2. The Main Results

Throughout this section, suppose that the process $(X_n, \theta_n)_{n\geq 1}$ follows the ASM recursion (S1)–(S3) described in Section 1. Before stating the first ergodicity result, consider the following condition on the regularity of a collection of sets.

Definition 1. Suppose that $\{A_i\}_{i\in I}$ with each $A_i \subset \mathbb{R}^d$ are such that there is a unique outer-pointing normal $n_i(x)$ for each x in the boundary ∂A_i. Then, $\{A_i\}_{i\in I}$ have *uniformly continuous normals* if for all $\epsilon > 0$ there is a $\delta > 0$ such that for any $i \in I$ it holds that $\|n_i(x) - n_i(y)\| \leq \epsilon$ for all $x, y \in \partial A_i$ such that $\|x - y\| \leq \delta$.

This definition essentially states that the boundaries ∂A_i must be regular enough to ensure that if one looks at ∂A_i at a small enough scale, it will look almost like a plane.

Theorem 2. *Assume π has a compact support $\mathbb{X} \subset \mathbb{R}^d$, and π is continuous, bounded, and bounded away from zero on $\mathbb{X}$. Moreover, assume that the set $\mathbb{X}$ has a uniformly continuous normal in the sense of Definition 1. Then, for any $0 < \alpha^* < 1/2$ and a bounded function f, the strong law of large numbers holds, i.e.*

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x \qquad \textit{almost surely.} \tag{1}$$

Theorem 2 follows as a special case of Theorem 22 in Section 5.

Let us consider next target distributions π with unbounded supports, satisfying the following conditions formulated in [20].

Assumption 3. The density π is bounded, bounded away from zero on compact sets, differentiable, and

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|^\rho} \cdot \nabla \log \pi(x) = -\infty \tag{2}$$

for some $\rho > 1$. Moreover, the contour normals satisfy

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|} \cdot \frac{\nabla \pi(x)}{\|\nabla \pi(x)\|} < 0. \tag{3}$$

This assumption is very near to the conditions introduced by Jarner and Hansen [12] to ensure the geometric ergodicity of a (non-adaptive) Metropolis algorithm, and considered by Andrieu and Moulines [1] in the context of adaptive MCMC. In particular, [1, 12] assume that π fulfils the contour regularity condition (3). Instead of (2), they assume a super-exponential decay on π,

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|} \cdot \nabla \log \pi(x) = -\infty$$

which is only slightly more general than (2). See [12] for examples and discussion on the conditions.

Theorem 4. *Suppose π fulfils Assumption 3 and there is a $t_0 > 0$ such that the contour sets $\{L_t\}_{0<t\le t_0}$ where $L_t := \{x \in \mathbb{R}^d : \pi(x) \ge t\}$ have uniformly continuous normals in the sense of Definition 1. Assume the function f is bounded on compact sets, and grows at most exponentially, i.e. $f(x) \le c_1 e^{c_2\|x\|}$ for all $\|x\| \ge 1$ with some constants $0 \le c_1, c_2 < \infty$. Then, for any $0 < \alpha^* < 1/2$, the strong law of large numbers* (1) *holds.*

Theorem 4 is a special case of Theorem 24 in Section 5.

Remark 5. For many practical target densities satisfying Assumption 3 the tail contours are (essentially) scaled copies of each other, in which case they have automatically uniformly continuous normals. This indicates that Theorem 4 is practically a counterpart of [20, Theorem 13] verifying the ergodicity of the Adaptive Metropolis algorithm.

Remark 6. The "safe" values for the desired acceptance rate stipulated by Theorems 2 and 4 are $0 < \alpha^* < 1/2$. This range is often sufficient, as the most commonly used values for a random-walk Metropolis algorithms are probably $\alpha^* = 0.234$ and $\alpha^* = 0.44$, and it has been suggested that values $0.1 \le \alpha^* \le 0.4$ should work well in most cases [8, 16, 18, 19].

Remark 7. The results below hold for the above algorithmic setting, but allow some modifications. Most importantly, one can use also a non-Gaussian proposal distribution q. In particular, the results hold for a heavy-tailed multivariate Student proposals. Moreover, one can employ a different weight sequence than $cn^{-\gamma}$ in (S3); the essential assumption is that the sum of square weights must be finite.

The rest of the paper is organised as follows. Section 3 describes a general adaptive MCMC scheme, and a generalised version of the above described ASM algorithm within that framework. Section 4 develops stability results for this process. In particular, Corollary 15 ensures the stability of the sequence θ_n with the assumptions of Theorem 2, and Proposition 18 controls the growth of θ_n when π fulfils the conditions of Theorem 4. Once the stability results are obtained, the ergodicity is verified in Section 5 by the results in [20].

3. Notations and Framework

The adaptive MCMC process considered here evolves in a measurable space $\mathbb{X} \times \mathbb{S}$, where $\mathbb{X}$ is the space of the "MCMC" chain $(X_n)_{n\ge1}$ and $\mathbb{S} = \mathbb{R}$ the space of the adaptation parameter chain $(S_n)_{n\ge1}$. The process starts at some given $X_1 \equiv x_1 \in \mathbb{X}$ and $S_1 \equiv s_1 \in \mathbb{S}$, and for $n \ge 1$, follows the recursion[1]

$$X_{n+1} = \begin{cases} Y_{n+1}, & \text{if } U_{n+1} \le \alpha_{S_n}(X_n, Y_{n+1}) \\ X_n, & \text{otherwise} \end{cases} \tag{4}$$

$$S_{n+1} = S_n + \eta_{n+1} H(S_n, X_n, Y_{n+1}) \tag{5}$$

[1] The recursion of (5) can be considered as Robbins-Monro stochastic approximation; see [1, 2, 4] and references therein.

where the acceptance probability $\alpha_s : \mathbb{X}\times\mathbb{X}\to[0,1]$ for each $s\in\mathbb{S}$, and $H:\mathbb{S}\times\mathbb{X}\times\mathbb{X}\to K_H$ is an adaptation function, with $K_H\subset\mathbb{R}$ compact. The random variables U_n and Y_n are assumed to be $\mathcal{F}_n$-measurable; U_{n+1} is independent on $\mathcal{F}_n$ and uniformly distributed on $[0,1]$, and Y_{n+1} depends on $\mathcal{F}_n$ only via X_n and S_n. Namely, Y_n are distributed by the proposal density q_s so that $\mathbb{P}(Y_{n+1}\in A\mid\mathcal{F}_n)=\int_A q_{S_n}(X_n,y)\mathrm{d}y$. The sequence of non-negative step sizes η_n decays to zero.

Hereafter, consider the following generalisation of the ASM algorithm of Section 1, formulated in the above framework. In that case, $\mathbb{X}\subset\mathbb{R}^d$ is the support of π and the family of proposal densities are defined as $q_s(x,y):=q_s(x-y)$ with

$$q_s(z):=[\phi(s)]^{-d}q([\phi(s)]^{-1}z)$$

where the template probability density q on $\mathbb{R}^d$ is symmetric, and the scaling function $\phi:\mathbb{R}\to(0,\infty)$ is increasing and surjective. To shed light on this definition, let Y be distributed according to q. Then, $\phi(s)Y$ is distributed according to q_s. In the context of the particular version of the algorithm described in Section 1, one has $\phi(s)=e^s$ and $S_n=\log\theta_n$. The acceptance probability is the Metropolis-Hastings ratio[2]

$$\alpha_s(x,y):=\alpha(x,y):=\min\left\{1,\frac{\pi(y)}{\pi(x)}\right\}.$$

The adaptation function H is defined as $H(s,x,y):=H(x,y):=\alpha(x,y)-\alpha^*$ where α^* is the constant desired acceptance rate, and the non-negative step sizes satisfy[3] $\sum_{k=1}^{\infty}\eta_k=\infty$ and $\sum_{k=1}^{\infty}\eta_k^2<\infty$.

Define the expected acceptance rate at $x\in\mathbb{X}$ with parameter $s\in\mathbb{S}$ as

$$A(x,s):=\int_{\mathbb{X}}\alpha(x,y)q_s(x-y)\mathrm{d}y.$$

Clearly, the adaptation rule decreases S_n whenever $A(X_n,S_n)<\alpha^*$, and vice versa. So, it is plausible that the algorithm would result in $S_n\to s^*$ such that $A(s^*)=\alpha^*$, where

$$A(s):=\int_{\mathbb{X}}A(x,s)\pi(x)\mathrm{d}x$$

is the expected acceptance rate over the target density π. In this paper, however, the convergence of S_n is not the main concern, but the stability of it, as it turns out to be crucial for the validity of the ASM algorithm.

The Metropolis transition kernel with a proposal density q_s is given as

$$P_s(x,A):=\mathbb{1}_A(x)\int_{\mathbb{R}^d}[1-\alpha(x,y)]q_s(x-y)\mathrm{d}y+\int_A\alpha(x,y)q_s(x-y)\mathrm{d}y \tag{6}$$

where $\mathbb{1}_A$ stands for the characteristic function of the set A. Using the kernels P_s, one can write (4) as $\mathbb{P}(X_{n+1}\in A\mid X_n=x,S_n=s)=P_s(x,A)$. As usual, integration

[2]Note that Y_{n+1} may lie outside $\mathbb{X}$, but $(X_n)_{n\geq1}\subset\mathbb{X}$ almost surely.

[3]The case $\sum_{k=1}^{\infty}\eta_k<\infty$ is not considered as it yields a trivially bounded S_n and prevents effective adaptation.

of a function f with respect to a kernel is denoted as

$$P_s f(x) := \int_{\mathbb{X}} f(y) P_s(x, \mathrm{d}y).$$

Let $V \geq 1$ be a function. The V-norm of a function f is defined as

$$\|f\|_V := \sup_x \frac{|f(x)|}{V(x)}.$$

The closed ball in $\mathbb{R}^d$ is written as $\overline{B}(x,r) := \{y \in \mathbb{R}^d : \|x-y\| \leq r\}$, and the distance of a point $x \in \mathbb{R}^d$ from the set $A \subset \mathbb{R}^d$ is denoted as $d(x,A) := \inf\{\|x-y\| : y \in A\}$.

4. Stability

This section develops stability results, starting with a simple theorem on the general process given in Section 3. This theorem is auxiliary for the present paper, but may have applications with other adaptive MCMC algorithms of similar type.

Theorem 8. *Suppose $(X_n, S_n)_{n\geq 1}$ follow the general recursions (4) and (5), and the step sizes satisfy $\sum_{n=1}^\infty \eta_n^2 < \infty$.*

(i) If there is a constant $a < \infty$ such that for all $n \geq 1$

$$\mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n\right] \leq 0 \qquad \text{whenever } S_n \geq a,$$

then $\limsup_{n\to\infty} S_n < \infty$ a.s.

(ii) If also $\sum_n \eta_n = \infty$, and there is a non-decreasing sequence of constants $(a_n)_{n\geq 1} \subset \mathbb{R}$ such that

$$\mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n\right] \leq b \qquad \text{whenever } S_n \geq a_n$$

for some $b < 0$, then $\limsup_{n\to\infty}(S_n - a_n) \leq 0$ a.s.

Proof. Let $W_n := H(S_{n-1}, X_{n-1}, Y_n)\mathbb{1}_{\{S_{n-1}\geq a\}}$ for $n \geq 2$, and define the martingale $(M_n, \mathcal{F}_n)_{n\geq 1}$ by setting $M_1 := 0$, and $M_n := \sum_{k=2}^n \mathrm{d}M_k$ for $n \geq 2$ with the differences $\mathrm{d}M_n := \eta_n(W_n - \mathbb{E}\left[W_n \mid \mathcal{F}_{n-1}\right])$. Clearly,

$$\sum_{k=2}^\infty \mathbb{E}\left[\mathrm{d}M_k^2 \mid \mathcal{F}_{k-1}\right] \leq 4c^2 \sum_{k=2}^\infty \eta_k^2 < \infty$$

where $c = \sup_{x\in K_H}|x|$. This implies that M_n converges to an a.s. finite limit M_∞ [e.g. 11, Theorem 2.15].

Let $(\tau_k)_{k\geq 1}$ be the exit times of S_n from $(-\infty, a)$, defined as $\tau_k := \inf\{n > \tau_{k-1} : S_n \geq a,\ S_{n-1} < a\}$ using the conventions $\tau_0 = 0$, $S_0 < a$, and $\inf\emptyset = \infty$. Define also the latest exit from $(-\infty, a)$ by $\sigma_n := \sup\{\tau_k : k \geq 1,\ \tau_k \leq n\}$. Whenever $S_n \geq a$, one can write $S_n = S_{\sigma_n} + (M_n - M_{\sigma_n}) + Z_{\sigma_n, n}$ where

$$Z_{m,n} := \sum_{k=m+1}^n \eta_n \mathbb{E}\left[W_n \mid \mathcal{F}_{n-1}\right] \leq 0$$

by assumption. In this case,

$$S_n \le S_{\sigma_n} + (M_n - M_{\sigma_n}) \le \max\{|S_1|, a + c\eta_{\sigma_n}\} + 2\sup_{k\ge 1}|M_k| \le C \tag{7}$$

where C is a.s. finite. If $S_n < a$ the claim is trivial and (i) holds.

Assume then (ii). If $S_n < a_n$ for all n greater than some $N_1(\omega) < \infty$, the claim is trivial. Suppose then that $S_n \ge a_n$ infinitely often. Define $(\tau_k)_{k\ge 1}$ as the exit times of S_n from $(-\infty, a_n)$ as above. The times τ_k are all a.s. finite in this case (and S_n returns to $(-\infty, a_n)$ infinitely often), for suppose the contrary; then the last exit times σ_n are bounded by some $\sigma_n \le \sigma < \infty$, and for $n \ge \sigma$ one may write

$$S_n = S_\sigma + (M_n - M_\sigma) + Z_{\sigma,n} \le C_\sigma + Z_{\sigma,n}$$

where M_n and $Z_{n,m}$ are defined as above, using $W_n := H(S_{n-1}, X_{n-1}, Y_n)\mathbb{1}_{\{S_{n-1}\ge a_{n-1}\}}$, and the random variable C_σ is a.s. finite as in (7). Now, $Z_{\sigma,n} \to -\infty$ a.s. as $n \to \infty$, so $S_n < a_n$ a.s. for sufficiently large n, which is a contradiction.

Fix an $\epsilon > 0$ and let $N_0 = N_0(\omega, \epsilon)$ be such that for all $n \ge N_0$, it holds that $c\eta_{\sigma_n} \le \epsilon/3$ and that $|M_k - M_\infty| \le \epsilon/3$ a.s. for all $k \ge \sigma_n$. The claim follows from the estimate

$$\begin{aligned} S_n &\le S_{\sigma_n} + (M_n - M_{\sigma_n}) = S_{\sigma_n - 1} + \eta_{\sigma_n} H(S_{\sigma_n-1}, X_{\sigma_n-1}, Y_{\sigma_n}) + (M_n - M_{\sigma_n}) \\ &\le a_{\sigma_n} + \epsilon/3 + |M_n - M_\infty| + |M_\infty - M_{\sigma_n}| \le a_n + \epsilon \end{aligned}$$

for all $n \ge N_0$. □

Remark 9. Theorem 8 generalises for an unbounded adaptation function H under suitable additional assumptions. For example, assuming

$$\limsup_{n\to\infty} |\eta_{n+1} H(S_n, X_n, Y_{n+1})| = 0 \qquad \text{and} \qquad \sum_{k=1}^{\infty} |\eta_{n+1} H(S_n, X_n, Y_{n+1})|^2 < \infty$$

hold almost surely, the proof applies with obvious changes. Moreover, the function H may depend additionally on U_{n+1} (or X_{n+1}).

Hereafter, consider the adaptive scaling Metropolis (ASM) algorithm described in Section 3. One can give simple conditions under which the result of Theorem 8 applies. This is due to the fact that one can write

$$\mathbb{E}[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n] = A(X_n, S_n) - \alpha^*,$$

so in light of Theorem 8, it is sufficient to find out when $A(x, s)$ is below or above α^*.

Proposition 10. *Assume π is supported on a compact set $\mathbb{X} \subset \mathbb{R}^d$ and $\alpha^* > 0$. Then, there is $b < 0$ and $a \in \mathbb{R}$ such that*

$$\mathbb{E}[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n] \le b \qquad \textit{whenever } S_n \ge a. \tag{8}$$

Proof. Without loss of generality, one can assume $0 \in \mathbb{X}$. Let $\epsilon > 0$ be sufficiently small so that $\int_{\overline{B}(0,\epsilon)} q(z)\mathrm{d}z \le \alpha^*/2$, and let a be sufficiently large so that $\phi(s) \ge$

$\mathrm{diam}(\mathbb{X})\epsilon^{-1}$ for all $s \geq a$. Then, for all $x \in \mathbb{X}$,

$$\int_{\mathbb{X}} \alpha(x,y) q_s(x-y)\mathrm{d}y \leq \int_{\mathbb{X}} [\phi(s)]^{-d} q([\phi(s)]^{-1}z)\mathrm{d}z = \int_{[\phi(s)]^{-1}\mathbb{X}} q(u)\mathrm{d}u$$
$$\leq \int_{\overline{B}(0,\epsilon)} q(u)\mathrm{d}u \leq \frac{\alpha^*}{2}$$

That is, (8) holds with $b = -\alpha^*/2 < 0$, whenever $s \geq a$. □

Before stating the next result bounding the conditional expectation to the opposite direction, let us consider a condition on the tails of π.

Assumption 11. There is a $\lambda > 0$ such that $L_\lambda := \{y \in \mathbb{R}^d : \pi(y) \geq \lambda\}$ is compact and π is continuous on L_λ. Moreover, the sets in the collection $\{L_t\}_{0<t\leq\lambda}$ have uniformly continuous normals in the sense of Definition 1.

Proposition 12. *Suppose the target density π satisfies Assumption 11. Then, for any $\alpha^* < 1/2$, there are $a \in \mathbb{R}$ and $b > 0$ such that*

$$\mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n\right] \geq b, \qquad \textit{whenever } S_n \leq a. \tag{9}$$

Before giving the proof of Proposition 12, let us outline the simple intuition behind it. For all s small enough, the mass of $q_s(\cdot)$ is essentially concentrated on a small ball $\overline{B}(0,\epsilon)$. If one looks the target π only on $\overline{B}(x,\epsilon)$, there are essentially two alternatives. The first one is that π is approximately constant on that small ball and $A(x,s) \approx 1$. The second alternative is that π decreases very rapidly to *one* direction, in which case the set $\{y : \pi(y) \geq \pi(x)\}$ looks like a half-space on the ball $\overline{B}(x,\epsilon)$, and $A(x,s) \gtrsim 1/2$.

Let us start with a lemma on this "half-space approximation."

Lemma 13. *Suppose that the sets $\{A_i\}_{i\in I}$ with $A_i \subset \mathbb{R}^d$ have uniformly continuous normals in the sense of Definition 1. Then, for any $\epsilon > 0$, there is a $\delta > 0$ such that for any $i \in I$, any $x \in A_i$ and any $0 \leq r \leq \delta$, there is a half-space T such that $\overline{B}(x,r) \cap T \subset \overline{B}(x,r) \cap A_i$, and the distance $d(x,T) \leq \epsilon r$.*

The claim is geometrically evident. The technical verification is given in Appendix A.

Proof of Proposition 12. Fix an $0 < \epsilon^* < 1$, and let $M \geq 1$ be sufficiently large so that

$$\int_{\overline{B}(0,\phi(s)M)} q_s(z)\mathrm{d}z = \int_{\overline{B}(0,M)} q(z)\mathrm{d}z \geq 1 - \epsilon^* \tag{10}$$

and for any plane P, it holds that

$$\int_{\{d(z,P)\leq\phi(s)M^{-1}\}} q_s(z)\mathrm{d}z = \int_{\{d(z,P)\leq M^{-1}\}} q(z)\mathrm{d}z \leq \epsilon^*. \tag{11}$$

By compactness of L_λ and positivity of π one can find $\delta_1 > 0$ such that for all $x, y \in L_\lambda$ with $\|x-y\| \le \delta_1$, it holds that $|\log\pi(x) - \log\pi(y)| \le \epsilon^*/e$ so that

$$1 - \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} = e^0 - e^{\min\{0,\log\pi(y)-\log\pi(x)\}} \le e|\log\pi(y) - \log\pi(x)| \le \epsilon^*.$$

Let $\delta_2 > 0$ be sufficiently small to satisfy Lemma 13 with the choice $\epsilon = M^{-2}$.

Choose a small enough $a \in \mathbb{R}$ so that $\phi(a)M \le \min\{\delta_1, \delta_2\}$. Let $s \le a$, denote $r_s := \phi(s)M$, and write for any $x \in L_\lambda$

$$\begin{aligned}\int_{\mathbb{X}} \alpha(x,y)q_s(x-y)\mathrm{d}y &\ge \int_{\overline{B}(x,r_s)\cap L_\lambda} \alpha(x,y)q_s(x-y)\mathrm{d}y \\ &\ge (1-\epsilon^*)\int_{\overline{B}(x,r_s)\cap L_\lambda} q_s(x-y)\mathrm{d}y\end{aligned}$$

since $r_s \le \delta_1$. Denote by T the half-space from Lemma 13, such that $\overline{B}(x,r_s)\cap T \subset \overline{B}(x,r_s)\cap L_\lambda$ and the distance $d(x,T) \le M^{-2}r_s$. One obtains

$$\begin{aligned}\int_{\mathbb{X}} \alpha(x,y)q_s(x-y)\mathrm{d}y &\ge (1-\epsilon^*)\int_{\overline{B}(x,r_s)\cap T} q_s(x-y)\mathrm{d}y \\ &\ge (1-\epsilon^*)\int_{\overline{B}(x,r_s)\cap\tilde{T}} q_s(x-y)\mathrm{d}y - \int_{\{d(y,P)\le M^{-2}r_s\}} q_s(x-y)\mathrm{d}y \\ &\ge \frac{1}{2}(1-\epsilon^*)^2 - \epsilon^*\end{aligned}$$

where $\tilde{T}$ is the half-space with the boundary plane P parallel to the boundary of T, and passing through x. The last inequality follows from (10) with the symmetry of q_s and (11), respectively. The same estimate holds for any $x \in L_t$ with $t > 0$.

To conclude,

$$\int_{\mathbb{X}} \alpha(x,y)q_s(x-y)\mathrm{d}y \ge \frac{1}{2} - \frac{1}{2}\left(\frac{1}{2} - \alpha^*\right)$$

for all $x \in \mathbb{X}$ and for any $\alpha^* < 1/2$ by selecting $\epsilon^* = \epsilon^*(\alpha^*) > 0$ to be sufficiently small, implying (9) with $b = (1/2 - \alpha^*)/2 > 0$. □

Remark 14. Assumption 11 in Proposition 12 is not minimal. For example, the sets L_t could have convex holes that do not have uniformly continuous normals, and one could still obtain Proposition 12.

As an easy corollary of the propositions above, one establishes the stability of the ASM process.

Corollary 15. *Assume the target density π is compactly supported, and satisfies Assumption 11. Then, for the ASM process $(X_n, S_n)_{n\ge1}$ with any $0 < \alpha^* < 1/2$, there are a.s. finite A_1 and A_2 such that*

$$A_1 \le S_n \le A_2 \tag{12}$$

for all $n \ge 1$.

Proof. The conditions of Propositions 10 and 12 are satisfied, so there are constants $-\infty < a_1 < a_2 < \infty$ and $b < 0$ such that

$$\begin{aligned} \mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n\right] &\le b \qquad \text{whenever} \qquad S_n \ge a_2, \\ \mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n\right] &\ge -b \qquad \text{whenever} \qquad S_n \le a_1. \end{aligned}$$

Theorem 8 applied to $-S_n$ and S_n guarantees that $a_1 \le \liminf_{n\to\infty} S_n$ and $\limsup_{n\to\infty} S_n \le a_2$, respectively, from which one obtains a.s. finite A_1 and A_2 for which (12) holds. □

The rest of this section considers targets π with an unbounded support. Under a suitably regular π, it is shown that the growth of S_n can be controlled. To start with, consider the following properties for the scaling function ϕ and the the template proposal distribution q.

Assumption 16. The scaling function ϕ is piecewise differentiable, and there are constants $h, c > 0$ and $\kappa \ge 1$ such that

$$\phi'(x + \xi) \le c \max\{1, \phi^\kappa(x)\}$$

for all $x \in \mathbb{R}$ and all $0 \le \xi \le h$.

Assumption 16 is not restrictive, and it clearly holds for any polynomially or exponentially increasing ϕ.

Assumption 17. The template proposal density q can be written as $q(z) = \hat{q}(\|\Sigma^{-1}z\|)$ where $\Sigma \in \mathbb{R}^{d\times d}$ is a symmetric and positive definite matrix, and $\hat{q} : [0,\infty) \to (0,\infty)$ is a bounded, decreasing, and differentiable function. Moreover, the derivative of $\hat{q}$ satisfies

(i) there is an $\epsilon^* > 0$ and $0 \le a < b < \infty$ such that for all $0 \le \epsilon \le \epsilon^*$, the following bounds hold

$$\begin{aligned} \hat{q}'(x) - 2\hat{q}'(x+\epsilon) &\ge c_1, \qquad \text{for all} \qquad a \le x \le b, \\ \int_0^\infty \min\{0, \hat{q}'(x) - 2\hat{q}'(x+\epsilon)\}\mathrm{d}x &\ge -c_2 e^{-c_3\epsilon^{-1}} \end{aligned}$$

with some constants $c_1, c_2, c_3 > 0$.

(ii) there are constants $c_4, c_5 > 0$ such that

$$\int_0^\infty r^d |\hat{q}'(\theta^{-1}r)|\mathrm{d}r \le c_4\theta_0^{c_5}$$

for all $0 < \theta < \theta_0$.

Assumption 17 stipulates that q is elliptically symmetric and the contours of q have main axes proportional to the eigenvalues of Σ. Moreover, the decay rate of q along any ray is determined by $\hat{q}$ satisfying the technical conditions (i) and (ii). Lemma 28 in Appendix B shows that Assumption 17 holds for Gaussian and Student distributions q.

The following estimate for the at most polynomial growth of $\phi(S_n)$ is crucial for the ergodicity result obtained in Theorem 24.

Proposition 18. *Suppose π satisfies Assumptions 3 and 11. Suppose also that the scaling function ϕ satisfies Assumption 16 and the template density q fulfils Assumption 17 (i). Then, for the ASM process $(X_n, S_n)_{n\geq 1}$ with $0 < \alpha^* < 1/2$, and for any $\beta > 0$, there is an a.s. positive $\Theta_1 = \Theta_1(\omega)$ and an a.s. finite $\Theta_2 = \Theta_2(\omega, \beta)$, such that*

$$\Theta_1 \leq \phi(S_n) \leq \Theta_2 n^\beta.$$

Before the proof, let us consider an estimate of $A(x, s)$ depending on both x and s.

Lemma 19. *Assume π satisfies Assumption 3. Then, for any $\epsilon > 0$, there is a constant $c = c(\epsilon) \geq 1$ such that $A(x, s) \leq \epsilon$ for all $\phi(s) \geq c \max\{1, \|x\|\}$.*

Proof. Let $r_1 \geq 1$ be sufficiently large so that for some $\gamma > 0$ it holds that $\frac{x}{\|x\|} \cdot \frac{\nabla\pi(x)}{\|\nabla\pi(x)\|} < -\gamma$ and $\frac{x}{\|x\|^\rho} \cdot \nabla \log \pi(x) < -\gamma$ for all $\|x\| \geq r_1$. Increase r_1, if necessary, so that for any $\|x\| \geq r_1$ one can write $L_{\pi(x)} = \{y : \pi(y) \geq \pi(x)\} = \{ru : u \in S^d, 0 \leq r \leq g(u)\}$ where $S^d := \{u \in \mathbb{R}^d : \|u\| = 1\}$ is the unit sphere and the function $g : S^d \to (0, \infty)$ parameterises the boundary of $L_{\pi(x)}$. Notice also that the contour normal condition implies the existence of an $M \geq 1$ such that $L_{\pi(x)} \subset \overline{B}(0, M\|x\|)$ for all $\|x\| \geq r_1$.

Write for $\|x\| \geq r_2 := Mr_1$

$$\begin{aligned} A(x,s) &= \int_{\mathbb{R}^d} \alpha(x,y) q_s(x-y)\mathrm{d}y \\ &\leq \int_{\{d(y,L_{\pi(x)})\leq\|x\|\}} q_s(x-y)\mathrm{d}y + \sup_{y\in\mathbb{R}^d} q_s(x-y) \int_{\{d(y,L_{\pi(x)})>\|x\|\}} \alpha(x,y)\mathrm{d}y. \end{aligned}$$

The first term can be estimated from above by

$$\int_{\overline{B}(0,M\|x\|+\|x\|)} q_s(x-y)\mathrm{d}y \leq \int_{\overline{B}(0,(M+2)\|x\|)} q_s(z)\mathrm{d}z = \int_{\overline{B}(0,r(s,x))} q(u)\mathrm{d}u \leq \frac{\epsilon}{2}$$

whenever $r(s, x) := [\phi(s)]^{-1}(M + 2)\|x\| \leq \epsilon^*$ for some small enough $\epsilon^* = \epsilon^*(\epsilon) > 0$, as in the proof of Proposition 10.

For the latter term, notice that $\sup_{y\in\mathbb{R}^d} q_s(x-y) = [\phi(s)]^{-d} \sup_{z\in\mathbb{R}^d} q(z) \leq c_1[\phi(s)]^{-d}$. The integral can be estimated by polar integration as

$$\int_{\{d(y,L_{\pi(x)})>\|x\|\}} \alpha(x,y)\mathrm{d}y \leq c_d \sup_{u\in S^d} \int_{r>g(u)+\|x\|}^{\infty} r^{d-1} e^{\log\pi(ru)-\log\pi(g(u)u)}\mathrm{d}r$$

where c_d is the surface measure of the sphere S^d. Since $\|x\| \geq r_2$, one has that $g(u) \geq r_1 \geq 1$, and from the gradient decay condition, one obtains that for $r > g(u)+1$

$$\begin{aligned} \log\pi(ru) - \log\pi(g(u)u) &= \int_{g(u)}^{r} \frac{tu}{\|tu\|} \cdot \nabla \log\pi(tu)\mathrm{d}t \leq -\gamma \int_{g(u)}^{r} t^{\rho-1}\mathrm{d}t \\ &\leq -\gamma g(u)^{\rho-1}[r - g(u)] \end{aligned}$$

from which

$$\int_{r>g(u)+\|x\|}^{\infty} r^{d-1} e^{\log \pi(ru) - \log \pi(g(u)u)} \mathrm{d}r \le \int_0^\infty e^{-\frac{\gamma w}{2}} \mathrm{d}w \sup_{r>g(u)+\|x\|} r^{d-1} e^{-\frac{\gamma}{2} g(u)^{\rho-1}[r-g(u)]}.$$

Consequently,

$$\int_{\{d(y,L_{\pi(x)})>\|x\|\}} \alpha(x,y)\mathrm{d}y \le \frac{2}{\gamma} \sup_{\tilde{g}\ge 1, \tilde{r}>1} \exp\Big[(d-1)\log(\tilde{g}+\tilde{r}) - \frac{\gamma}{2}\tilde{g}^{\rho-1}\tilde{r}\Big] \le c_2$$

with a constant $c_2 > 0$ whenever $\|x\| \ge r_2$.

To sum up, there is a $c_3 > 0$ such that for $\|x\| \ge r_2$ and

$$\phi(s) \ge c_3 \max\{1, \|x\|\} \ge \max\left\{1,\, c_1 c_2 \frac{2}{\epsilon},\, \frac{(M+2)\|x\|}{\epsilon^*}\right\}$$

it holds that $A(x,s) \le \epsilon$. For any $\|x\| < r_2$ there is a $r_2 \le \|x_0\| \le M r_2$ such that $\pi(x_0) \le \pi(x)$. Consequently, $\alpha(x,y) \le \alpha(x_0,y)$ for all $y \in \mathbb{R}^d$ and therefore $A(x,s) \le A(x_0,s)$. To conclude, $A(x,s) \le \epsilon$ for all $\phi(s) \ge M c_3 \max\{1,\|x\|\}$. □

Having Lemma 19 and the lower bound from Proposition 12, the proof of Proposition 18 can be obtained fairly easily using the a growth condition established in [20].

Proof of Proposition 18. Proposition 12 applied with Theorem 8 for $-S_n$ gives an a.s. finite A_1 such that $A_1 \le S_n$. Since $\phi > 0$ is increasing, the variable $\Theta_1 := \phi(A_1)$ is a.s. positive, showing the lower bound.

To check the polynomial growth condition for $\phi(S_n)$, it is first verified that $\|X_n\|$ grows at most polynomially. Fix an $\epsilon > 0$, and let $\theta_1 = \theta_1(\epsilon) > 0$ and $a_1 = a_1(\epsilon) \in \mathbb{R}$ be such that $\theta_1 = \phi(a_1)$, and that $\mathbb{P}(B_1) \ge 1-\epsilon$, with $B_1 := \{\Theta_1 \ge \theta_1\} = \{A_1 \ge a_1\}$. Let $V(x) := c_\pi \pi^{-1/2}(x)$, where the constant $c_\pi := [\sup_x \pi(x)]^{1/2}$ ensures that $V \ge 1$. Proposition 26 in Appendix B shows that the drift inequality

$$P_s V(x) \le V(x) + b \tag{13}$$

holds for all $\phi(s) \ge \theta_1 > 0$ with some $b = b(\theta_1) < \infty$. Construct an auxiliary process $(X_n', S_n')_{n\ge 1}$ coinciding with $(X_n, S_n)_{n\ge1}$ in B_1 by setting $(X_n', S_n') = (X_{\tau_n}, S_{\tau_n})$ where the stopping times τ_n are defined as

$$\tau_n := \begin{cases} n, & \text{if } S_k \ge \theta_1 \text{ for all } 1 \le k \le n \\ \inf\{1 \le k \le n-1 : S_{k+1} < \theta_1\}, & \text{otherwise.} \end{cases}$$

Having the inequality (13), set $\beta' = \kappa\beta$ and use Proposition 10 of [20] to obtain the bound $\|X_n'\| \le \Theta_\epsilon n^{\beta'}$ for some a.s. finite Θ_ϵ. The $\epsilon > 0$ was arbitrary, so one can let $\epsilon \to 0$ and obtain an a.s. finite Θ such that $\|X_n\| \le \Theta n^{\beta'}$. Applying Lemma 19, one obtains that $A(X_n, S_n) \le \alpha^*/2$ whenever $\phi(S_n) \ge \Theta' n^{\beta'}$ with $\Theta' := c_1 \max\{1, \Theta\}$.

Fix again an $\epsilon > 0$ and let $\theta_2 = \theta_2(\epsilon) < \infty$ be such that $\mathbb{P}(B_2) \ge 1-\epsilon$ where $B_2 := \{\Theta' \le \theta_2\}$. Construct an auxiliary process $(X_n', S_n')_{n\ge1}$ coinciding with $(X_n, S_n)_{n\ge1}$

in B_2 by stopping the process if $S_k > \theta_2 k^{\beta'}$ as in the construction above. Theorem 8 ensures that

$$\limsup_{n\to\infty}[S_n' - \tilde{a}_n] \leq 0$$

where $\tilde{a}_n$ are defined so that $\phi(\tilde{a}_n) = \theta_2 n^{\beta'}$. That is, $S_n' \leq \tilde{a}_n + E_n$ with $E_n \to 0$ almost surely. Consider Assumption 16 and take N_0 so large that $E_n < h$ for all $n \geq N_0$. Then, $\phi(x+h) = \phi(x) + h\phi'(x+\xi)$ for some $0 \leq \xi \leq h$, and hence $\phi(x+h) \leq c_2 \max\{1, \phi(x)^\kappa\}$. For $n \geq N_0$, one has

$$\phi(S_n') \leq \phi(\tilde{a}_n + E_n) \leq c_2 \max\{1, \phi(\tilde{a}_n)^\kappa\} = c_2 \max\{1, \theta_2^\kappa n^{\kappa\beta'}\} \leq \theta_2' n^\beta.$$

Summing up, there is an a.s. finite Θ_2' such that

$$\phi(S_n') \leq \Theta_2' n^\beta$$

on B_2. Finally, letting $\epsilon \to 0$, one can find an a.s. finite Θ_2 such that $\phi(S_n) \leq \Theta_2 n^\beta$. □

Remark 20. It is possible to obtain Corollary 15 and Proposition 18 when using the ASM algorithm within some other adaptation framework. For example, ASM can be combined with the Adaptive Metropolis algorithm as suggested in [5] and [3]. In particular, one could assume that there is another ($\mathcal{F}_n$-measurable) parameter $\tilde{S}_n$ in addition to S_n, so that $Y_{n+1} \sim q_{S_n,\tilde{S}_n}(X_n, \cdot)$ with

$$q_{s,\tilde{s}}(x,y) := q_{s,\tilde{s}}(x-y) := [\phi(s)]^{-d} q_{\tilde{s}}([\phi(s)]^{-1}(x-y))$$

where $\{q_{\tilde{s}}\}_{\tilde{s}\in\tilde{\mathbb{S}}}$ is a suitably "uniform" family of probability densities. If there are integrable $q^+, q^- \geq 0$ such that $q^-(z) \leq q_{\tilde{s}}(z) \leq q^+(z)$ for all $\tilde{s} \in \tilde{\mathbb{S}}$ and with $\int_{\mathbb{R}^d} q^-(z)\mathrm{d}z > 0$ as suggested in [7], then Corollary 15 can be easily verified to hold. Similarly, Proposition 18 can be shown to hold if additionally all $\{q_{\tilde{s}}\}_{\tilde{s}\in\tilde{\mathbb{S}}}$ satisfy Assumption 17 (i).

5. Ergodicity

Section 4 described conditions under which the ASM algorithm was shown to be stable, or to have a controlled growth. This section formulates strong laws of large numbers for the ASM process, following the technique introduced in [20]. For this purpose, an alternative theoretical adaptation recursion is formulated, applying a sequence of restriction sets $K_1 \subset K_2 \subset \cdots \subset K_n \subset \mathbb{S}$.

Assume $(\tilde{X}_n, \tilde{S}_n)_{n\geq 1}$ follow the adaptation framework described in Section 3. Assume $\tilde{S}_1 \equiv \tilde{s}_1 \in K_1$, and instead of (5) let $(\tilde{S}_n)_{n\geq 1}$ follow the "truncated" recursion

$$\tilde{S}_{n+1} = \sigma_{n+1}\left(\tilde{S}_n, \eta_{n+1} H(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_{n+1})\right) \tag{14}$$

where the restriction functions $\sigma_n : \mathbb{S} \times \mathbb{S} \to \mathbb{S}$ are defined as

$$\sigma_n(s,s') := \begin{cases} s+s', & \text{if } s+s' \in K_n \\ s, & \text{otherwise.} \end{cases}$$

That is, σ_n ensures that $\tilde{S}_n \in K_n$ for all $n \geq 1$. Observe that if $K_n = \mathbb{S}$ for all $n \geq 1$, then the truncated recursion (14) reduces to the original adaptation recursion (5).

Before stating an ergodicity result for this truncated chain, four technical assumptions are listed, which must hold for some constants $c \geq 1$ and $\epsilon \geq 0$.

(A1) For each $s \in \mathbb{S}$, the transition probability P_s has π as the unique invariant distribution.

(A2) For each $n \geq 1$, the following uniform drift and minorisation conditions hold for all $s \in K_n$

$$
\begin{aligned}
P_s V(x) &\leq \lambda_n V(x) + b_n \mathbb{1}_{C_n}(x), && \forall x \in \mathbb{X} \\
P_s(x, A) &\geq \delta_n \nu_s(A), && \forall x \in C_n, \forall A \subset \mathbb{X}
\end{aligned}
$$

where $C_n \subset \mathbb{X}$ is a subset (a minorisation set), $V : \mathbb{X} \to [1, \infty)$ is a drift function such that $\sup_{x \in C_n} V(x) \leq b_n$, and ν_s is a probability measure on $\mathbb{X}$, concentrated on C_n. Furthermore, the constants $\lambda_n \in (0,1)$ and $b_n \in (0, \infty)$ are increasing, and $\delta_n \in (0,1]$ is decreasing with respect to n, and they are polynomially bounded so that

$$\max\{(1-\lambda_n)^{-1}, \delta_n^{-1}, b_n\} \leq c n^\epsilon.$$

(A3) For all $n \geq 1$ and any $r \in (0,1]$, there is $c' = c'(r) \geq 1$ such that for all $s, s' \in K_n$,

$$\|P_s f - P_{s'} f\|_{V^r} \leq c' n^\epsilon \|f\|_{V^r} |s - s'|.$$

(A4) The inequality $|H(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_{n+1})| \leq c n^\epsilon$ holds almost surely.

Theorem 21. *Assume (A1)–(A4) hold and let f be a function with $\|f\|_{V^\beta} < \infty$ for some $\beta \in (0,1)$. Assume $\epsilon < \kappa_*^{-1} \min\{1/2, 1-\beta\}$, where $\kappa_* \geq 1$ is an independent constant, and that $\sum_{k=1}^\infty k^{\kappa_* \epsilon - 1} \eta_k < \infty$. Then,*

$$\frac{1}{n} \sum_{k=1}^{n} f(\tilde{X}_k) \xrightarrow{n \to \infty} \int_{\mathbb{X}} f(x) \pi(x) \mathrm{d}x \quad \textit{almost surely.} \tag{15}$$

Proof. This theorem is a straightforward modification of Theorem 2 in [20]. In particular, the assumption (A4) here is slightly simpler than assumption (A4) in [20], and the changes required for the proof are obvious. □

The following first main result considers the case of compactly supported π.

Theorem 22. *Suppose π has a compact support $\mathbb{X} \subset \mathbb{R}^d$, and π is continuous, bounded, and bounded away from zero on $\mathbb{X}$. Moreover, assume that the set $\mathbb{X}$ has a uniformly continuous normal in the sense of Definition 1, and the template proposal density q satisfies Assumption 17. Then, for the ASM process $(X_n, S_n)_{n \geq 1}$ with any $0 < \alpha^* < 1/2$ and a bounded function f, the strong law of large numbers holds, i.e.*

$$\frac{1}{n} \sum_{k=1}^{n} f(X_k) \xrightarrow{n \to \infty} \int_{\mathbb{R}^d} f(x) \pi(x) \mathrm{d}x \tag{16}$$

almost surely.

Proof. Corollary 15 ensures that for any $\delta > 0$, there are $-\infty < a_1^{(\delta)} < a_2^{(\delta)} < \infty$ such that $\mathbb{P}(B^{(\delta)}) \geq 1 - \delta$, where

$$B^{(\delta)} := \{\forall n \geq 1, \quad a_1^{(\delta)} \leq S_n \leq a_2^{(\delta)}\}.$$

Set $K_n^{(\delta)} := K^{(\delta)} := [a_1^{(\delta)}, a_2^{(\delta)}]$ for all $n \geq 1$, and construct the truncated process $(\tilde{X}_n^{(\delta)}, \tilde{S}_n^{(\delta)})$ using these restriction sets in (14). Define $\theta_1^{(\delta)} := \phi(a_1^{(\delta)}) > 0$ and $\theta_2^{(\delta)} := \phi(a_2^{(\delta)}) < \infty$.

Let us next verify above assumptions with some $c \geq 1$ and $\epsilon = 0$, and for $V \equiv 1$. The assumption (A1) holds by construction of the Metropolis process. For (A2), take $C_n := \mathbb{X}$ for all $n \geq 1$, and notice that $P_s V(x) = 1$ for all $x \in \mathbb{X}$ and $s \in \mathbb{S}$. By Assumption 17 (i) one can estimate for all $s \in K^{(\delta)}$ and all $x \in \mathbb{X}$,

$$\begin{aligned} P_s(x, A) &\geq \int_A \alpha(x,y) q_s(x-y)\mathrm{d}y \geq \left(\inf_{x,y\in\mathbb{X},\, s\in K^{(\delta)}} q_s(x-y) \right) \int_A \frac{\pi(y)}{\sup_{z\in\mathbb{X}} \pi(z)} \mathrm{d}y \\ &\geq \theta_2^{-d} \left(\inf_{|z|\leq \mathrm{diam}(\mathbb{X})} \hat{q}(\|\theta_1^{-1}\Sigma^{-1} z\|) \right) c_1 \nu_s(A) \geq \delta \nu_s(A) \end{aligned}$$

with a $\delta > 0$, where $\nu_s(A) := \nu(A) := c_1^{-1} \int_A \frac{\pi(y)}{\sup_{z\in\mathbb{X}}\pi(z)} \mathrm{d}y$ and $c_1 > 0$ chosen so that $\nu(\mathbb{X}) = 1$. Assumption 16 ensures that the derivative of ϕ is bounded on the compact set $K_n^{(\delta)}$. Therefore, the Frobenius norm $\|\phi(s)\Sigma - \phi(s')\Sigma\| \leq c_2|s - s'|$ with some $c_2(\delta) > 0$, and Proposition 27 in Appendix B implies (A3). Finally, it holds that $|H(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_{n+1})| \leq c$, implying (A4).

All (A1)–(A4) hold and $\sum_{k=1}^{\infty} k^{-1}\eta_k \leq (\sum_{k=1}^{\infty} k^{-2})^{1/2} (\sum_{k=1}^{\infty} \eta_k^2)^{1/2} < \infty$, so Theorem 21 yields a strong law of large numbers for the truncated process $\tilde{X}_n^{(\delta)}$. Since $(\tilde{X}_n^{(\delta)})_{n\geq 1}$ coincides with the original ASM process $(X_n)_{n\geq 1}$ in $B^{(\delta)}$, the strong law of large numbers applies for $X_n(\omega)$ with almost every $\omega \in B^{(\delta)}$. Since $\delta > 0$ is arbitrary, (16) holds almost surely. □

Remark 23. Theorem 21 is actually a version of Theorem 2 of [20], which is a modification of Proposition 6 in [1]. Theorem 22 could be obtained also using other techniques, in particular, the mixingale approach described in [6, 10], or the coupling technique of [15] (resulting in a weak law of large numbers). These other techniques do not, however, apply directly to Theorem 24 below, where Theorem 21 is applied in full strength.

Finally, the second main result considers target densities π with unbounded support.

Theorem 24. *Suppose π satisfies Assumptions 3 and 11. Assume the function f is bounded on compact sets, and grows at most exponentially, i.e. $f(x) \leq c_1 e^{c_2\|x\|}$ for all $\|x\| \geq 1$ with some constants $0 \leq c_1, c_2 < \infty$. Then, for the ASM process $(X_n, S_n)_{n\geq 1}$ with any $0 < \alpha^* < 1/2$, the strong law of large numbers* (16) *holds.*

Proof. Proposition 18 ensures that for any $\epsilon' > 0$ there are a.s. positive Θ_1 and a.s. finite Θ_2 such that

$$\Theta_1 \le \phi(S_n) \le \Theta_2 n^{\epsilon'}. \tag{17}$$

Now, similarly as in the proof of Theorem 22, for any $\delta > 0$, one can find $0 < \theta_1^{(\delta)} \le \theta_2^{(\delta)} < \infty$ such that

$$\mathbb{P}(\forall n \ge 1 : \theta_1^{(\delta)} \le S_n \le \theta_2^{(\delta)} n^{\epsilon'}) \ge 1 - \delta \tag{18}$$

and construct $(\tilde{X}_n^{(\delta)}, \tilde{S}_n^{(\delta)})_{n\ge 1}$ using the restriction sets $K_n^{(\delta)} := [a_1^{(\delta)}, a_2^{(n,\delta)}]$, where $\phi(a_1^{(\delta)}) = \theta_1^{(\delta)}$ and $\phi(a_2^{(\delta,n)}) = \theta_2^{(\delta)} n^{\epsilon'}$.

Let $V(x) := c_V \pi^{-1/2}(x)$ with $c_V := (\sup_x \pi(x))^{1/2}$. The assumptions (A1) and (A4) hold as verified in the proof of Theorem 22. Proposition 26 in Appendix B with the fact $\det(\theta\Sigma) = \theta^d \det(\Sigma)$ yields (A2) with $\epsilon = d\epsilon'$. Assumption 16 ensures that $\phi'(s) \le c\phi^\kappa(s)$, from which $|\phi(s) - \phi(s')| \le c(\theta_2^{(\delta)} n^{\epsilon'})^\kappa |s - s'| \le \tilde{c} n^{\kappa\epsilon'} |s - s'|$. Now, Proposition 27 in Appendix B shows (A3) with $\epsilon = c_2\kappa\epsilon'$. To conclude, the assumptions (A1)–(A4) hold with constants (c, ϵ), where $\epsilon = \epsilon(\delta, \epsilon') > 0$ can be selected to be arbitrarily small, and $c = c(\delta, \epsilon) < \infty$.

In particular, one can let ϵ be sufficiently small to ensure that $\kappa_*\epsilon < 1/3$ and $\rho - \epsilon > 1$. Then, $\sum_{k=1}^\infty k^{\kappa_*\epsilon - 1}\eta_k < \infty$ as in the proof of Theorem 22 and $V(x) \ge c_3 e^{\|x\|}$. Theorem 21 ensures that the strong law of large numbers holds in the set (18), and a.s. by letting $\delta \to 0$. □

Acknowledgements

The author wishes to thank Prof. Eero Saksman for comments significantly improving the presentation of the paper.

References

[1] C. Andrieu and É. Moulines. On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.*, 16(3):1462–1505, 2006.
[2] C. Andrieu and C. P. Robert. Controlled MCMC for optimal sampling. Technical Report Ceremade 0125, Université Paris Dauphine, 2001.
[3] C. Andrieu and J. Thoms. A tutorial on adaptive MCMC. *Statist. Comput.*, 18 (4):343–373, Dec. 2008.
[4] C. Andrieu, É. Moulines, and P. Priouret. Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.*, 44(1):283–312, 2005.
[5] Y. Atchadé and G. Fort. Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. URL `http://arxiv.org/abs/0807.2952`. Preprint, July 2008.
[6] Y. F. Atchadé and J. S. Rosenthal. On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, 11(5):815–828, 2005.

[7] Y. Bai, G. O. Roberts, and J. S. Rosenthal. On the containment condition for adaptive Markov chain Monte Carlo algorithms. URL `http://probability.ca/jeff/research.html`. Preprint, July 2008.

[8] M. Bédard. Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234. *Stochastic Process. Appl.*, 118(12):2198–2222, 2008.

[9] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, Boca Raton, Florida, 1998. ISBN 0-412-05551-1.

[10] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001. ISSN 1350-7265.

[11] P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980. ISBN 0-12-319350-8.

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

[13] E. Nummelin. MC's for MCMC'ists. *International Statistical Review*, 70(2): 215–240, 2002.

[14] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 1999. ISBN 0-387-98707-X.

[15] G. O. Roberts and J. S. Rosenthal. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.*, 44(2):458–475, 2007.

[16] G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. Preprint, 2006.

[17] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

[18] G. O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367, 2001.

[19] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. 7(1):110–120, 1997.

[20] E. Saksman and M. Vihola. On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. URL `http://arxiv.org/abs/0806.2933`. Preprint, June 2008.

Appendix A. Half-Space Approximation

Proof of Lemma 13. Fix an $\epsilon' > 0$. By the uniform smoothness of $\{\partial A_i\}_{i\in I}$, one can let $\delta > 0$ be so small that $\|n_i(y) - n_i(z)\| \le \epsilon'$ for all $i \in I$ and $y, z \in \partial A_i$ with $\|y - z\| \le 2\delta$.

Fix an $i \in I$, an $x \in A_i$, and a $r \in [0, \delta]$. If $\overline{B}(x, r) \setminus A_i = \emptyset$, one can let T be any half-space passing through x. Suppose next $\overline{B}(x, r) \setminus A_i \neq \emptyset$, and let $y \in \overline{B}(x, r) \cap \partial A_i$. Consider the open cones

$$C_- := \{y + z : n_i(y) \cdot z < -\epsilon' \|z\|\}$$
$$C_+ := \{y + z : n_i(y) \cdot z > \epsilon' \|z\|\}$$

illustrated in Figure 1. We shall verify that $\overline{B}(y, 2\delta) \cap C_- \subset \overline{B}(y, 2\delta) \cap A_i$ and

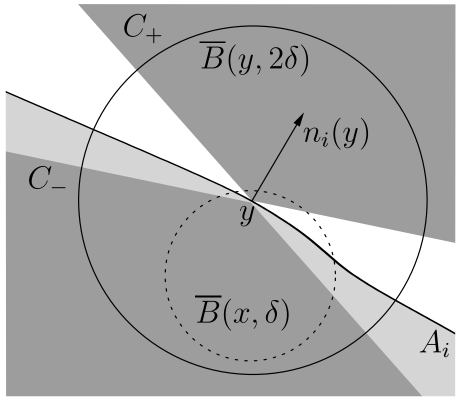

FIGURE 1. Illustration of the half-space approximation. The set A_i is shown in light grey, and the cones C_- and C_+ in dark grey.

$\overline{B}(y, 2\delta) \cap C_+ \subset \overline{B}(y, 2\delta) \setminus A_i$.

Namely, let $u \in \overline{B}(y, 2\delta) \cap C_-$ and write $u = y + z$. Suppose that $u \notin A_i$ and define $t_0 := \inf\{t \in [0, 1] : y + tz \notin A_i\}$. Let $u_0 := y + t_0 z$ and notice that $u_0 \in \overline{B}(y, 2\delta) \cap \partial A_i$. Moreover, the line segment $y + tz$ with $t \in [0, 1]$ passes through ∂A_i at u_0, and therefore $n_i(u_0) \cdot z \geq 0$. On the other hand,

$$n_i(u_0) \cdot \frac{z}{\|z\|} = (n_i(u_0) - n_i(y)) \cdot \frac{z}{\|z\|} + n_i(y) \cdot \frac{z}{\|z\|} < \|n_i(u_0) - n_i(y)\| - \epsilon' < 0,$$

which is a contradiction, implying That is, $C_- \cap \overline{B}(y, 2\delta) \subset A_i \cap \overline{B}(y, 2\delta)$. The case with C_+ is similar.

Let $T := \{y - 2\epsilon' r n_i(y) + z : z \cdot n_i(y) < 0\}$. It holds that $\overline{B}(y, 2r) \cap T \subset \overline{B}(y, 2r) \cap C_-$ since taking $y + w \in \overline{B}(y, 2r) \cap T$ one has $n_i(y) \cdot w < -2\epsilon' r \leq -\epsilon' \|w\|$. On the other hand, $\overline{B}(y, 2r) \cap C_- \subset \overline{B}(y, 2r) \cap A_i$ and $\overline{B}(x, r) \subset \overline{B}(y, 2r)$, so $\overline{B}(x, r) \cap T \subset \overline{B}(x, r) \cap A_i$. Clearly, $d(y, T) = 2\epsilon' r$, and since $x \notin C_+$ one has $n_i(y) \cdot (x - y) \leq \epsilon' \|x - y\| \leq \epsilon' r$. To conclude, $d(x, T) \leq 3\epsilon' r$, and taking $\epsilon' = \epsilon/3$ yields the claim. □

APPENDIX B. SIMULTANEOUS PROPERTIES FOR METROPOLIS KERNELS

Let us define the following generalisation of Assumption 17.

Assumption 25. Let $\mathcal{C}_d \subset \mathbb{R}^{d \times d}$ stand for the symmetric and positive definite matrices. Suppose $\mathcal{P} \subset \mathcal{C}_d$ and $\{q_s\}_{s \in \mathcal{P}}$ is a family of probability densities defined through

$$q_s(z) := |\det(s)|^{-1} \hat{q}(\|s^{-1} z\|), \tag{19}$$

where $\hat{q} : [0, \infty) \to (0, \infty)$ is a bounded, decreasing, and differentiable function. Moreover, suppose that there is a $\kappa > 0$ such that the eigenvalues of each $s \in \mathcal{P}$ are bounded from below by κ.

Proposition 26. *Suppose π satisfies Assumption 3 and the family $\{q_s\}_{s \in \mathcal{P}}$ satisfies Assumption 25 with some $\kappa > 0$. Moreover, suppose that $\hat{q}$ in* (19) *satisfies Assumption 17 (i). Let P_s be the Metropolis transition probability defined in* (6) *and using the* *proposal density q_s. Then, there exists a compact set $C \subset \mathbb{R}^d$, a probability measure ν on C, and a constant $b \in [0, \infty)$ such that for any $s \in \mathcal{P}$*

$$P_s V(x) \le \lambda_s V(x) + b \mathbb{1}_C(x), \qquad \forall x \in \mathbb{X} \tag{20}$$

$$P_s(x, B) \ge \delta_s \nu(B) \qquad \forall x \in C, \forall B \subset \mathbb{X} \tag{21}$$

where $V(x) := c_V \pi^{-1/2}(x) \ge 1$ with $c_V := (\sup_x \pi(x))^{1/2}$ and the constants $\lambda_s, \delta_s \in (0,1)$ satisfy the bound

$$(1-\lambda_s)^{-1} \vee \delta_s^{-1} \le c |\det(s)|^{-1}$$

for some constant $c \ge 1$.

Proof. Proposition 26 is a generalisation of [20, Proposition 18] considering Gaussian densities q_s. We shall describe the changes that are needed in the proof of [20, Proposition 18].

Let $s \in \mathcal{P}$. For a non-negative function f, one can write by Fubini's theorem

$$\begin{aligned} \int_{\mathbb{R}^d} f(z+x) q_s(z) \mathrm{d}z &= |\det(s)|^{-1} \int_0^{\hat{q}(0)} \int_{\{\hat{q}(\|[s^{-1}z\|) \ge t\}} f(z+x) \mathrm{d}z \mathrm{d}t \\ &= -|\det(s)|^{-1} \int_0^\infty \int_{E_u} f(y) \mathrm{d}y \hat{q}'(u) \mathrm{d}u \end{aligned}$$

where the substitution $t = \hat{q}(u)$ was used, and $E_u := \{x + z : \|s^{-1}z\| \le u\}$. One has $\|s^{-1}z\| \le \kappa^{-1}\|z\|$, and thus $E_u \supset \overline{B}(x, u\kappa)$. Assumption 17 (i) for the derivative $\hat{q}'$ corresponds to the estimate obtained in [20, Lemma 17] for a Gaussian family, i.e. $\hat{q} = e^{-x^2/2}$.

These facts are enough to complete the proof of [20, Proposition 18] to yield the claim. □

Proposition 27. *Suppose the family $\{q_s\}_{s\in\mathcal{P}}$ satisfies Assumption 25 with some $\kappa > 0$. Suppose, in addition, that $\hat{q}$ fulfils Assumption 17 (ii), and either*

(i) $V \equiv 1$, or

(ii) π satisfies Assumption 3 and $V(x) := c_V \pi^{-1/2}(x) \ge 1$ with $c_V := (\sup_x \pi(x))^{1/2}$.

Then, there are constants $c_1, c_2 > 0$ such that for the Metropolis transition probability P_s given in (6), it holds that

$$\|P_s f - P_{s'} f\|_{V^r} \le c_1 \max\{\|s\|, \|s'\|\}^{c_2} \|f\|_{V^r} \|s - s'\| \tag{22}$$

for all $s, s' \in \mathcal{P}$ and $r \in [0,1]$. The matrix norm above is the Frobenius norm defined as $\|a\| := \sqrt{\mathrm{tr}(a^T a)}$.

Proof. Consider first (i). From the definition of the Metropolis kernel (6), one obtains

$$\sup_x |P_s f(x) - P_{s'} f(x)| \le 2 \sup_x |f(x)| \int_{\mathbb{X}} |q_s(x) - q_{s'}(x)| \mathrm{d}x.$$

For (ii), Proposition 12 of [1] shows that for any $r \in [0,1]$ it holds that

$$\|P_s f - P_{s'} f\|_{V^r} \le 2 \|f\|_{V^r} \int_{\mathbb{X}} |q_s(x) - q_{s'}(x)| \mathrm{d}x$$

so it is sufficient to consider only the total variation of the proposal distributions.

As in [10] and [1], one can write

$$\int_{\mathbb{X}} |q_s(x) - q_{s'}(x)|\mathrm{d}x = \int_{\mathbb{X}} \left| \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} q_{s_t}(x)\mathrm{d}t \right| \mathrm{d}x.$$

where $s_t := s' + t(s - s')$. Let us compute

$$\frac{\mathrm{d}}{\mathrm{d}t} q_{s_t}(x) = -\operatorname{tr}\left(s_t^{-1}(s - s')\right) q_{s_t}(x) + |\det(s_t)|^{-1} \hat{q}'(\|s_t^{-1}x\|) \frac{\mathrm{d}}{\mathrm{d}t}\|s_t^{-1}x\|$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\|s_t^{-1}x\| = -\left(\frac{s_t^{-1}x}{\|s_t^{-1}x\|}\right)^T s_t^{-1}(s - s')s_t^{-1}x.$$

Since $s - s'$ and s_t are symmetric and s_t positive definite, it holds that $|\operatorname{tr}\left(s_t^{-1}(s - s')\right)| \le \operatorname{tr}(s_t^{-1}) \max_{1\le i\le d} |\lambda_i| \le \operatorname{tr}(s_t^{-1})\|s - s'\|$ where λ_i are the eigenvalues of $s - s'$. Since the Frobenius norm is sub-multiplicative,

$$\begin{aligned}
&\int_{\mathbb{X}} |q_s(x) - q_{s'}(x)|\mathrm{d}x \\
&\quad \le \sup_{t\in[0,1]} \left(\operatorname{tr}(s_t^{-1}) + |\det(s_t)|^{-1} \|s_t^{-1}\|^2 \int_{\mathbb{X}} \|x\| \left|\hat{q}'(\|s_t^{-1}x\|)\right| \mathrm{d}x \right) \|s - s'\| \\
&\quad \le \left(d\kappa^{-1} + \kappa^{-d} d\kappa^{-2} c_d \sup_{\|u\|=1,\, t\in[0,1]} \int_0^\infty r^d |\hat{q}'(r\|s_t^{-1}u\|)|\mathrm{d}r \right) \|s - s'\| \\
&\quad \le c_1 \lambda^{c_2} \|s - s'\|
\end{aligned}$$

by polar integration and Assumption 17 (ii), where λ is the maximum eigenvalue of s and s'. Clearly, $\lambda \le \max\{\|s\|, \|s'\|\}$ concluding the proof. □

Lemma 28. *Suppose the template proposal density q is given as $q(z) = c\hat{q}(\|\Sigma^{-1}z\|)$ where $c > 0$ is a constant and $\Sigma \subset \mathbb{R}^{d\times d}$ is a symmetric and positive definite matrix, and*

(i) $\hat{q}(x) = e^{-x^2/2}$, or
(ii) $\hat{q}(x) = (1 + x^2)^{-d/2-\gamma}$ for some $\gamma > 0$.

That is, q is a (multivariate) Gaussian or Student distribution, respectively. Then, q satisfies Assumption 17.

Proof. Consider first (i). Assumption 17 (i) is implied by [20, Lemma 17]. For Assumption 17 (ii), let $\theta_0 > 0$.

$$\int_0^\infty r^d |\hat{q}'(\theta^{-1}r)|\mathrm{d}r = \theta^{-1} \int_0^\infty r^{d+1} e^{-\frac{r^2}{2\theta^2}} \mathrm{d}r = c_d \theta^{d+1} \int_0^\infty u^{\frac{d}{2}} e^{-u} \mathrm{d}u \le c\theta_0^{d+1}$$

for all $0 < \theta < \theta_0$ with $c = c(d) > 0$.

Assume then that $\hat{q}$ has the form (ii) and fix an $\epsilon > 0$. By the mean value theorem, denoting $c_1 := d + 2\gamma$ and $\alpha := d/2 + \gamma + 1$, one can write for some $\epsilon' \in [0, \epsilon]$

$$\begin{aligned}
\hat{q}'(x) - 2\hat{q}'(x+\epsilon) &\geq c_1 x \left(\frac{2}{(1+(x+\epsilon)^2)^\alpha} - \frac{1}{(1+x^2)^\alpha} \right) \\
&= c_1 x \left(\frac{1}{(1+(x+\epsilon)^2)^\alpha} - \frac{2\alpha\epsilon(x+\epsilon')}{(1+(x+\epsilon')^2)^{\alpha+1}} \right) \\
&\geq \frac{c_1 x}{(1+(x+\epsilon)^2)^\alpha} \left(1 - 2\alpha\epsilon \left(\frac{1+(x+\epsilon)^2}{1+(x+\epsilon')^2} \right)^\alpha \right) > 0
\end{aligned}$$

for all $x > 0$, whenever $\epsilon > 0$ is sufficiently small, showing Assumption 17 (i). Let us compute

$$\int_0^\infty r^d |\hat{q}'(\theta^{-1} r)| \mathrm{d}r = c_1 \int_0^\infty \frac{r^{d+1}\theta^{-1}\mathrm{d}r}{(1+\theta^{-2}r^2)^\alpha} = c_1 \theta^{d+1} \int_0^\infty \frac{u^{d+1}\mathrm{d}u}{(1+u^2)^\alpha} \leq c_3 \theta_0^{d+1}$$

for any $0 < \theta < \theta_0$, where the constant $c_3 = c_3(d, \gamma) < \infty$. □

Matti Vihola, Department of Mathematics and Statistics, University of Jyväskylä, P.O.Box 35 (MaD), FI-40014 University of Jyväskylä, Finland

E-mail address: matti.vihola@iki.fi

URL: http://iki.fi/mvihola/