The Annals of Applied Probability
2010, Vol. 20, No. 6, 2178–2203
DOI: 10.1214/10-AAP682

ON THE ERGODICITY OF THE ADAPTIVE METROPOLIS ALGORITHM ON UNBOUNDED DOMAINS

BY EERO SAKSMAN[1] AND MATTI VIHOLA[2]

University of Helsinki and University of Jyväskylä

This paper describes sufficient conditions to ensure the correct ergodicity of the Adaptive Metropolis (AM) algorithm of Haario, Saksman and Tamminen [*Bernoulli* **7** (2001) 223–242] for target distributions with a noncompact support. The conditions ensuring a strong law of large numbers require that the tails of the target density decay super-exponentially and have regular contours. The result is based on the ergodicity of an auxiliary process that is sequentially constrained to feasible adaptation sets, independent estimates of the growth rate of the AM chain and the corresponding geometric drift constants. The ergodicity result of the constrained process is obtained through a modification of the approach due to Andrieu and Moulines [*Ann. Appl. Probab.* **16** (2006) 1462–1505].

1. Introduction.

The Markov chain Monte Carlo (MCMC) method, first proposed by [11], is a commonly used device for numerical approximation of integrals of the type

$$\pi(f) = \int f(x)\pi(x)\,dx,$$

where π is a probability density function. Intuitively, the method is based on producing a sample $(X_k)_{k=1}^n$ of random variables from the distribution π defines. The integral $\pi(f)$ is approximated with the average $I_n := n^{-1}\sum_{k=1}^n f(X_k)$. In particular, the random variables $(X_k)_{k=1}^n$ are a realization of a Markov chain, constructed so that the chain has π as the unique invariant distribution.

One of the most commonly applied constructions of such a chain in $\mathbb{R}^d$ is to let $X_0 \equiv x_0$ with some fixed point $x_0 \in \mathbb{R}^d$, and recursively for $n \geq 1$:

1. simulate $Y_n = X_{n-1} + U_n$, where U_n is an independent random variable distributed according to some symmetric proposal distribution q, for example, a zero-mean Gaussian, and

Received June 2008; revised February 2009.

[1]Supported by the Academy of Finland, Projects 113826 and 118765, and by the Finnish Centre of Excellence in Analysis and Dynamics Research.

[2]Supported by the Academy of Finland, Projects 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.

AMS 2000 subject classifications. Primary 65C05; secondary 65C40, 60J27, 93E15, 93E35.

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

2. with probability $\min\{1, \pi(Y_n)/\pi(X_{n-1})\}$, 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 is often efficient enough, even in a relatively complex and high-dimensional situation, provided that the proposal distribution q is selected properly. Finding a good proposal for a particular problem can, however, be a difficult task.

Recently, there has been a number of publications describing different adaptation techniques aiming to find a good proposal automatically [1, 3, 5, 9, 13] (see also the review article [4]). It has been a common practice to perform trial runs, and determine the proposal from the outcome. The recently proposed methods are different in that they adapt on-the-fly, continuously during the estimation run. In this paper, we focus on the forerunner of these methods, the Adaptive Metropolis (AM) algorithm [9], which is a random-walk Metropolis sampler with a Gaussian proposal q_v having a covariance v. The proposal covariance v is updated continuously during the run, according to the history of the chain. In general, such an adaptation may, if carelessly implemented, destroy the correct ergodicity properties, that is, that I_n does not converge to $\pi(f)$ as $n \to \infty$ (see, e.g., [13]). For practical considerations of the AM algorithm, the reader may consult [8, 14].

In the original paper [9] presenting the AM algorithm, the first ergodicity result for such adaptive algorithms was obtained. More precisely, a strong law of large numbers was proved for bounded functionals, when the algorithm is run on a compact subset of $\mathbb{R}^d$. After that, several authors have obtained more general conditions under which an adaptive MCMC process preserves the correct ergodicity properties. Andrieu and Robert [3] established the connection between adaptive MCMC and stochastic approximation, and proposed a general framework for adaptation. Atchadé and Rosenthal [5] developed further the technique of [9]. Andrieu and Moulines [1] made important progress by generalizing the Poisson equation and martingale approximation techniques to the adaptive setting. They proved the ergodicity and a central limit theorem for a class of adaptive MCMC schemes. Roberts and Rosenthal [13] use an interesting approach based on coupling to show a weak law of large numbers. However, in the case of AM, all the techniques essentially assume that the adapted parameter is constrained to a predefined compact set, or do not present concrete verifiable conditions. The only result to overcome this assumption is the one by Andrieu and Moulines [1]. Their result, however, requires a modification of the algorithm, including additional re-projections back to some fixed compact set.

This paper describes sufficient conditions under which the AM algorithm preserves the correct ergodicity properties, and $I_n \to \pi(f)$ almost surely as $n \to \infty$ for any function f that is bounded on compact sets and grows at most exponentially as $\|x\| \to \infty$. Our main result (Theorem 10) holds for the original AM process (without re-projections) having a target distribution supported on $\mathbb{R}^d$. Essentially, the target density π must have asymptotically lighter tails than $\pi(x) = ce^{-\|x\|^p}$ for some $p > 1$, and for large enough $\|x\|$, the sets $A_x = \{y \in \mathbb{R}^d : \pi(y) \geq \pi(x)\}$ must have uniformly regular contours. Our assumptions are very close to the well-known conditions proposed by Jarner and Hansen [10] to ensure the geometric convergence of a (nonadaptive) Metropolis process. By the techniques of this paper, one may also establish a central limit theorem (see Theorem 18).

The ergodicity results for the AM process rely on three main contributions. First, in Section 2, we describe an adaptive MCMC framework, in which the adaptation parameter is constrained at each time to a feasible adaptation set. In Section 3, we prove a strong law of large numbers for such a process, through a modification of the technique of Andrieu and Moulines [1]. Second, we propose an independent estimate for the growth rate of a process satisfying a general drift condition in Section 4. Third, in Section 5, we provide an estimate for constants of geometric drift for a symmetric random-walk Metropolis process, when the target distribution has super-exponentially decaying tails with regular contours.

The paper is essentially self-contained, and assumes little background knowledge. Only the basic martingale theory is needed to follow the argument, with the exception of Theorem 19 by Meyn and Tweedie [12], restated in Appendix A. Even though we consider only the AM algorithm, our techniques apply also to many other adaptive MCMC schemes of similar type.

2. General framework and notation. We consider an adaptive Markov chain Monte Carlo (MCMC) chain evolving in space $\mathbb{X} \times \mathbb{S}$, where $\mathbb{X}$ is the state space of the "MCMC" chain $(X_n)_{n\geq 0}$ and the adaptation parameter $(S_n)_{n\geq 0}$ evolves in $\mathbb{S} \subset \overline{\mathbb{S}}$, where $\overline{\mathbb{S}}$ is a separable normed vector space. We assume an underlying probability space $(\Omega, \mathcal{F}_\Omega, \mathbb{P})$, and denote the expectation with respect to $\mathbb{P}$ by $\mathbb{E}$. The natural filtration of the chain is denoted with $\mathcal{F} := (\mathcal{F}_k)_{k\geq 0} \subset \mathcal{F}_\Omega$ where $\mathcal{F}_k := \sigma(X_j, S_j : 0 \leq j \leq k)$. We also assume that we are given an increasing sequence $K_0 \subset K_1 \subset \cdots \subset K_n \subset \mathbb{S}$ of subsets of the adaptation parameter space $\mathbb{S}$. The random variables $(X_n, S_n)_{n\geq 0}$ form a stochastic chain, starting from $S_0 \equiv s_0 \in K_0 \subset \mathbb{S}$ and $X_0 \equiv x_0 \in \mathbb{X}$, and for $n \geq 0$, satisfying the following recursion:

$$X_{n+1} \sim P_{S_n}(X_n, \cdot), \tag{1}$$

$$S_{n+1} = \sigma_{n+1}(S_n, \eta_{n+1} H(S_n, X_{n+1})), \tag{2}$$

where P_s is a transition probability for each $s \in \mathbb{S}$, $H : \mathbb{S} \times \mathbb{X} \to \overline{\mathbb{S}}$ is an adaptation function, and $(\eta_n)_{n\geq 1}$ is a decreasing sequence of adaptation step sizes $\eta_n \in (0, 1)$. The functions $\sigma_n : \mathbb{S} \times \overline{\mathbb{S}} \to \mathbb{S}$ are defined as

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

Thus, σ_n ensures that S_n lies in K_n for each $n \geq 0$. The recursion (2) can also be considered as constrained Robbins–Monro stochastic approximation (see [1, 2] and references therein).

Let $V : \mathbb{X} \to [1, \infty)$ be a function. We define a V-norm of a function f as

$$\|f\|_V := \sup_{x \in \mathbb{X}} \frac{|f(x)|}{V(x)}.$$

As usual, we denote the integration of a function f with respect to a (signed) measure μ as $\mu(f) := \int f(x)\mu(dx)$, and define $Pf(x) := \int f(y)P(x, dy)$ for a transition probability P. The V-norm of a signed measure is defined as

$$\|\mu\|_V := \sup_{|f| \le V} |\mu(f)|.$$

The indicator function of a set A is denoted as $\mathbb{1}_A(x)$ and equals one if $x \in A$ and zero otherwise. In addition, we use the notation $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$.

Finally, we define the following regularity property for a family of functions $\{f_s\}_{s \in \mathbb{S}}$.

DEFINITION 1. Suppose $V : \mathbb{X} \to [1, \infty)$. Given an increasing sequence of subsets $K_n \subset \mathbb{S}$, $n \ge 1$, we say that a family of functions $\{f_s\}_{s \in \mathbb{S}}$, with $f_s : \mathbb{X} \to \mathbb{R}$, is (K_n, V)-polynomially Lipschitz with constants $c \ge 1$, $\varepsilon \ge 0$, if for all $s, s' \in K_n$, we have

$$\|f_s\|_V \le cn^{\varepsilon} \quad \text{and} \quad \|f_s - f_{s'}\|_V \le cn^{\varepsilon}|s - s'|.$$

3. Ergodicity of sequentially constrained adaptive MCMC. This section contains general ergodicity results for a sequentially constrained process defined in Section 2. These results can be seen auxiliary to our results on Adaptive Metropolis in Section 5, but may be applied to other adaptive MCMC methods as well.

Suppose that the adaptation algorithm has the form given in (1) and (2), and the following assumptions are satisfied for some $c \ge 1$ and $\varepsilon \ge 0$.

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

(A2) For each $n \ge 1$, the following uniform drift and minorization condition holds for all $s \in K_n$:

$$P_s V(x) \le \lambda_n V(x) + b_n \mathbb{1}_{C_n}(x) \qquad \forall x \in \mathbb{X}, \tag{3}$$

$$P_s(x, A) \ge \delta_n \nu_s(A) \qquad \forall x \in C_n, \forall A \subset \mathbb{X}, \tag{4}$$

where $C_n \subset \mathbb{X}$ is a subset (a minorization set), $V : \mathbb{X} \to [1, \infty)$ is a drift function such that $\sup_{x \in C_n} V(x) \le b_n$, 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

$$(1 - \lambda_n)^{-1} \vee \delta_n^{-1} \vee b_n \le cn^{\varepsilon}.$$

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

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

(A4) There is a $\beta \in [0, 1/2]$ such that for all $n \geq 1$, $s \in K_n$ and $x \in \mathbb{X}$

$$|H(s, x)| \leq c n^{\varepsilon} V^{\beta}(x).$$

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

$$\frac{1}{n} \sum_{k=1}^{n} f(X_k) \xrightarrow{n \to \infty} \pi(f) \qquad \text{almost surely.} \tag{5}$$

The proof of Theorem 2 is postponed to the end of this section. We start by the following lemma, whose proof is given in Appendix A. It shows that if we have polynomially worse bounds for drift and minorization constants, then the speed of geometric convergence can get only polynomially worse.

LEMMA 3. *Suppose (A2) holds. Then, one has for $r \in (0, 1]$ that for all $s \in K_n$ and $k \geq 1$,*

$$\|P_s^k(x, \cdot) - \pi(\cdot)\|_{V^r} \leq V^r(x) L_n \rho_n^k$$

with bound

$$L_n \vee (1 - \rho_n)^{-1} \leq c_2 n^{\kappa_2 \varepsilon},$$

where $\kappa_2 > 0$ is an independent constant, and $c_2 = c_2(c, r) \geq 1$.

Observe that the statement in Lemma 3 entails that any function $\|f\|_V < \infty$ is integrable with respect to the measures π and $P_s^k(x, \cdot)$, for all $x \in \mathbb{X}$, $k \geq 1$ and $s \in \bigcup_{n \geq 0} K_n$. The next three results are modified from Proposition 3, Lemma 5 and Proposition 6 of [1], respectively. The first one bounds the regularity of the solutions $\hat{f}_s$ of the Poisson equation

$$\hat{f}_s - P_s \hat{f}_s = f_s - \pi(f_s) \tag{6}$$

for a polynomially Lipschitz family of functions.

PROPOSITION 4. *Suppose that (A1)–(A3) hold, and the family of functions $\{f_s\}_{s \in \mathbb{S}}$ is (K_n, V^r)-polynomially Lipschitz with constants (c, ε), for some $r \in (0, 1]$. There is an independent constant $\kappa_3 > 0$ and a constant $c_3 = c_3(c, c', r) \geq 1$, such that:*

(i) *The family* $\{P_s f_s\}_{s\in\mathbb{S}}$ *is* (K_n, V^r)*-polynomially Lipschitz with constants* $(c_3, \kappa_3\varepsilon)$.

(ii) *Define, for any* $s \in \mathbb{S}$, *the function*

$$\hat{f}_s := \sum_{k=0}^{\infty}[P_s^k f_s - \pi(f_s)]. \tag{7}$$

Then, $\hat{f}_s$ *solves the Poisson equation* (6), *and the families* $\{\hat{f}_s\}_{s\in\mathbb{S}}$ *and* $\{P_s\hat{f}_s\}_{s\in\mathbb{S}}$ *are* (K_n, V^r)*-polynomially Lipschitz with constants* $(c_3, \kappa_3\varepsilon)$. *In other words,*

$$\|\hat{f}_s\|_{V^r} + \|P_s\hat{f}_s\|_{V^r} \le c_3 n^{\kappa_3\varepsilon}, \tag{8}$$

$$\|\hat{f}_s - \hat{f}_{s'}\|_{V^r} + \|P_s\hat{f}_s - P_{s'}\hat{f}_{s'}\|_{V^r} \le c_3 n^{\kappa_3\varepsilon}|s - s'| \tag{9}$$

for all $s, s' \in K_n$.

PROOF. Throughout the proof, suppose $s, s' \in K_n$.

The part (i) follows easily from Lemma 3, since

$$\begin{aligned}\|P_s f_s\|_{V^r} &\le \|P_s f_s - \pi(f_s)\|_{V^r} + |\pi(f_s)| \le [c_2 n^{\kappa_2\varepsilon} + \pi(V^r)]\|f_s\|_{V^r},\\ \|P_s f_s - P_{s'} f_{s'}\|_{V^r} &\le \|(P_s - P_{s'})f_s\|_{V^r} + \|P_{s'}(f_s - f_{s'})\|_{V^r}\\ &\le c' n^{\varepsilon}\|f_s\|_{V^r}|s - s'| + \tilde{c} n^{\kappa_2\varepsilon}\|f_s - f_{s'}\|_{V^r} \le \tilde{c} n^{(\kappa_2+1)\varepsilon}|s - s'|.\end{aligned}$$

Consider then (ii). Estimate (8) follows by the definition of $\hat{f}_s$ and Lemma 3,

$$\begin{aligned}\|\hat{f}_s\|_{V^r} &\le \sum_{k=0}^{\infty}\|P_s^k f_s - \pi(f_s)\|_{V^r} \le L_n\|f_s\|_{V^r}\sum_{k=0}^{\infty}\rho_n^k\\ &= \frac{L_n}{1-\rho_n}\|f_s\|_{V^r} \le (c_2 n^{\kappa_2\varepsilon})^2 c n^{\varepsilon} = c_2^2 c n^{(2\kappa_2+1)\varepsilon}.\end{aligned}$$

The above bound clearly applies also to $\|P_s\hat{f}_s\|_{V^r}$, and the convergence implies that $\hat{f}_s$ solves (6).

For (9), define an auxiliary transition probability by setting $\Pi(x, A) := \pi(A)$, and write

$$P_s^k f - P_{s'}^k f = \sum_{j=0}^{k-1}(P_s^j - \Pi)(P_s - P_{s'})[P_{s'}^{k-j-1} f - \pi(f)]$$

since $\pi P_s = \pi$ for all s. By Lemma 3 and assumption (A3), we have for all $s, s' \in K_n$ and $j \ge 0$

$$\begin{aligned}&\|(P_s^j - \Pi)(P_s - P_{s'})[P_{s'}^{k-j-1} f - \pi(f)]\|_{V^r}\\ &\quad\le L_n\rho_n^j\|(P_s - P_{s'})[P_{s'}^{k-j-1} f - \pi(f)]\|_{V^r}\\ &\quad\le L_n\rho_n^j c' n^{\varepsilon}|s - s'|\|P_{s'}^{k-j-1} f - \pi(f)\|_{V^r}\\ &\quad\le L_n^2\rho_n^{k-1},\end{aligned}$$

which gives that

$$\|P_s^k f - P_{s'}^k f\|_{V^r} \le k L_n^2 \rho_n^{k-1} c' n^{\varepsilon} |s - s'| \|f\|_{V^r}. \tag{10}$$

Write then

$$\hat{f}_s - \hat{f}_{s'} = \sum_{k=0}^{\infty} [P_s^k f_s - P_{s'}^k f_s] - \sum_{k=0}^{\infty} [P_{s'}^k (f_{s'} - f_s) - \pi(f_{s'} - f_s)].$$

By Lemma 3 and estimate (10) we have

$$\begin{aligned} \|\hat{f}_s - \hat{f}_{s'}\|_{V^r} &\le L_n^2 c' n^{\varepsilon} |s - s'| \left(\sum_{k=0}^{\infty} k \rho_n^{k-1}\right) \|f_s\|_{V^r} + L_n \left(\sum_{k=0}^{\infty} \rho_n^k\right) \|f_{s'} - f_s\|_{V^r} \\ &\le [L_n^2 c' n^{\varepsilon} (1 - \rho_n)^{-2} c n^{\varepsilon} + L_n (1 - \rho_n)^{-1} c n^{\varepsilon}] |s - s'| \\ &\le [(c_2 n^{\kappa_2 \varepsilon})^2 c' n^{\varepsilon} (c_2 n^{\kappa_2 \varepsilon})^2 c n^{\varepsilon} + (c_2 n^{\kappa_2 \varepsilon})(c_2 n^{\kappa_2 \varepsilon}) c n^{\varepsilon}] |s - s'| \\ &\le 2 c_2^4 c' c n^{(4\kappa_2 + 2)\varepsilon} |s - s'|. \end{aligned}$$

The same bound applies, with a similar argument, to $P_s \hat{f}_s - P_{s'} \hat{f}_{s'}$. $\square$

LEMMA 5. *Assume that* (A2) *holds. Then, for all* $r \in [0, 1]$, *any sequence* $(a_n)_{n \ge 1}$ *of positive numbers, and* $(x_0, s_0) \in \mathbb{X} \times K_0$, *we have that*

$$\mathbb{E}[V^r(X_k)] \le c_4^r k^{2r\varepsilon} V^r(x_0), \tag{11}$$

$$\mathbb{E}\Big[\max_{m \le j \le k} (a_j V(X_j))^r\Big] \le c_4^r \left(\sum_{j=m}^{k} a_j j^{2\varepsilon}\right)^r V^r(x_0), \tag{12}$$

where the constant c_4 *depends only on* c.

PROOF. For $(x_0, s_0) \in \mathbb{X} \times K_0$ and $k \ge 1$, we can apply the drift inequality (3) and the monotonicity of λ_k and b_k to obtain

$$\begin{aligned} \mathbb{E}[V(X_k)] &= \mathbb{E}[\mathbb{E}[V(X_k) | \mathcal{F}_{k-1}]] = \mathbb{E}[P_{S_{k-1}} V(X_{k-1})] \\ &\le \lambda_k \mathbb{E}[V(X_{k-1})] + b_k \le \cdots \le \lambda_k^k V(x_0) + b_k \sum_{j=0}^{k-1} \lambda_k^j \qquad (13) \\ &\le \left(1 + b_k \sum_{j=0}^{\infty} \lambda_k^j\right) V(x_0) \le (1 + c^2 k^{2\varepsilon}) V(x_0) \le c_4 k^{2\varepsilon} V(x_0). \end{aligned}$$

This estimate with Jensen's inequality yields for $r \in [0, 1]$ that

$$\mathbb{E}[V^r(X_k)] \le (\mathbb{E}[V(X_k)])^r \le c_4^r k^{2r\varepsilon} V^r(x_0).$$

Similarly, we have

$$\begin{aligned}\mathbb{E}\Big[\max_{m\le j\le k}(a_j V(X_j))^r\Big] &\le \Big(\mathbb{E}\Big[\max_{m\le j\le k} a_j V(X_j)\Big]\Big)^r\\ &\le \left(\sum_{j=m}^{k} a_j \mathbb{E}[V(X_j)]\right)^r\\ &\le c_4^r\left(\sum_{j=m}^{k} a_j j^{2\varepsilon}\right)^r V^r(x_0)\end{aligned}$$

by Jensen's inequality and estimate (13). □

Assume that $\{f_s\}_{s\in\mathbb{S}}$ is a regular enough family of functions. Consider the following decomposition, which is one of the key observations in [1],

$$\sum_{j=1}^{k}[f_{S_j}(X_j) - \pi(f_{S_j})] = M_k + R_k^{(1)} + R_k^{(2)}, \tag{14}$$

where $(M_k)_{k\ge1}$ is a martingale with respect to $\mathcal{F}$, and $(R_k^{(1)})_{k\ge1}$ and $(R_k^{(2)})_{k\ge1}$ are "residual" sequences, given by

$$\begin{aligned}M_k &:= \sum_{j=1}^{k}[\hat{f}_{S_{j-1}}(X_j) - P_{S_{j-1}}\hat{f}_{S_{j-1}}(X_{j-1})],\\ R_k^{(1)} &:= \sum_{j=1}^{k}[\hat{f}_{S_j}(X_j) - \hat{f}_{S_{j-1}}(X_j)],\\ R_k^{(2)} &:= P_{S_0}\hat{f}_{S_0}(X_0) - P_{S_k}\hat{f}_{S_k}(X_k).\end{aligned}$$

Recall that $\hat{f}_s$ solves the Poisson equation (6). The following proposition controls the fluctuations of these terms individually.

PROPOSITION 6. *Assume (A1)–(A4) hold,* $(x_0, s_0) \in \mathbb{X}\times K_0$ *and let* $\{f_s\}_{s\in\mathbb{S}}$ *be* (K_n, V^α)*-polynomially Lipschitz with constants* (c,ε) *for some* $\alpha\in(0,1-\beta)$*. Then, for any* $p\in(1,(\alpha+\beta)^{-1}]$*, for all* $\delta>0$ *and* $\xi>\alpha$*, there is a* $c_* = c_*(c,p,\alpha,\beta,\xi)\ge 1$*, such that for all* $n\ge1$*,*

$$\mathbb{P}\left[\sup_{k\ge n}\frac{|M_k|}{k}\ge\delta\right] \le c_*\delta^{-p} n^{p\varepsilon_*-(p/2)\wedge(p-1)} V^{\alpha p}(x_0), \tag{15}$$

$$\mathbb{P}\left[\sup_{k\ge n}\frac{|R_k^{(1)}|}{k^\xi}\ge\delta\right] \le c_*\delta^{-p}\left(\sum_{j=1}^{\infty}(j\vee n)^{\varepsilon_*-\xi}\eta_j\right)^p V^{(\alpha+\beta)p}(x_0), \tag{16}$$

$$\mathbb{P}\left[\sup_{k\ge n}\frac{|R_k^{(2)}|}{k^\xi}\ge\delta\right] \le c_*\delta^{-p} n^{p\varepsilon_*-(\xi-\alpha)p} V^{\alpha p}(x_0), \tag{17}$$

whenever $\varepsilon > 0$ *is small enough to ensure that* $\varepsilon_* := \kappa_* \varepsilon < [\frac{1}{2} \wedge (1 - \frac{1}{p}) \wedge (\xi - \alpha)]$, *where* $\kappa_* \geq 1$ *is an independent constant.*

PROOF. In this proof, $\tilde{c}$ is a constant that can take different values at each appearance. By Proposition 4, we have that $\|\hat{f}_s\|_{V^\alpha} + \|P_s \hat{f}_s\|_{V^\alpha} \leq c_3 \ell^{\kappa_3 \varepsilon}$ for all $s \in K_\ell$. Since $\alpha p \in [0, 1]$, we can bound the martingale differences $dM_\ell := M_\ell - M_{\ell-1}$ for $\ell \geq 1$ as follows:

$$
\begin{aligned}
\mathbb{E}|dM_\ell|^p &= \mathbb{E}|\hat{f}_{S_{\ell-1}}(X_\ell) - P_{S_{\ell-1}} \hat{f}_{S_{\ell-1}}(X_{\ell-1})|^p \\
&\leq \mathbb{E}\big|\|\hat{f}_{S_{\ell-1}}\|_{V^\alpha} V^\alpha(X_\ell) + \|P_{S_{\ell-1}} \hat{f}_{S_{\ell-1}}\|_{V^\alpha} V^\alpha(X_{\ell-1})\big|^p \\
&\leq 2^p (c_3 \ell^{\kappa_3 \varepsilon})^p \big(\mathbb{E}[V^{\alpha p}(X_\ell)] + \mathbb{E}[V^{\alpha p}(X_{\ell-1})]\big) \\
&\leq 2^{p+1} c_3^p c_4^{\alpha p} \ell^{p \kappa_3 \varepsilon} \ell^{2\alpha p \varepsilon} V^{\alpha p}(x_0) \leq \tilde{c} \ell^{(\kappa_3 + 2\alpha) p \varepsilon} V^{\alpha p}(x_0)
\end{aligned}
\tag{18}
$$

by (11) of Lemma 5. For $p \geq 2$, we have, by Burkholder and Minkowski's inequalities,

$$
\begin{aligned}
\mathbb{E}|M_k|^p &\leq c_p \mathbb{E}\left[\sum_{\ell=1}^k |dM_\ell|^2\right]^{p/2} \leq c_p \left[\sum_{\ell=1}^k (\mathbb{E}|dM_\ell|^p)^{2/p}\right]^{p/2} \\
&\leq \tilde{c} k^{(\kappa_3 + 2\alpha) p \varepsilon + p/2} V^{\alpha p}(x_0),
\end{aligned}
$$

where the constant c_p depends only on p. For $1 < p \leq 2$, the estimate (18) yields, by Burkholder's inequality,

$$
\begin{aligned}
\mathbb{E}|M_k|^p &\leq c_p \mathbb{E}\left[\sum_{\ell=1}^k (|dM_\ell|^p)^{2/p}\right]^{p/2} \leq c_p \sum_{\ell=1}^k \mathbb{E}|dM_\ell|^p \\
&\leq \tilde{c} k^{(\kappa_3 + 2\alpha) p \varepsilon + 1} V^{\alpha p}(x_0).
\end{aligned}
$$

The two cases combined give that

$$
\mathbb{E}|M_k|^p \leq \tilde{c} k^{(\kappa_3 + 2\alpha) p \varepsilon + (p/2) \vee 1} V^{\alpha p}(x_0). \tag{19}
$$

Now, by Corollary 21 of Birnbaum and Marshall's inequality in Appendix B,

$$
\begin{aligned}
\mathbb{P}\left[\max_{n \leq k \leq m} \frac{|M_k|}{k} \geq \delta\right] &\leq \delta^{-p}\left[m^{-p} \mathbb{E}|M_m|^p + \sum_{k=n}^{m-1} \big(k^{-p} - (k+1)^{-p}\big) \mathbb{E}|M_k|^p\right] \\
&\leq \delta^{-p}\left[m^{-p} \mathbb{E}|M_m|^p + p \sum_{k=n}^{m-1} k^{-p-1} \mathbb{E}|M_k|^p\right]
\end{aligned}
$$

for all $m \geq n$. By letting $\kappa_* := \kappa_3 + 3$, we have from (19)

$$
m^{-p} \mathbb{E}|M_m|^p \leq \tilde{c} m^{p(\kappa_* \varepsilon + (1/2) \vee (1/p) - 1)} \xrightarrow{m \to \infty} 0,
$$

since $\kappa_*\varepsilon + (1/2) \vee (1/p) < 1$. Now, (15) follows by

$$\mathbb{P}\left[\sup_{k\geq n} \frac{|M_k|}{k} \geq \delta\right] \leq \tilde{c}\delta^{-p}\left[\sum_{k=n}^{\infty} k^{(\kappa_3+2\alpha)p\varepsilon+(p/2)\vee 1-p-1}\right] V^{\alpha p}(x_0)$$
$$\leq \tilde{c}\delta^{-p} n^{p\kappa_*\varepsilon-(p/2)\wedge(p-1)} V^{\alpha p}(x_0)$$

since we have that $p\kappa_*\varepsilon - (p/2) \wedge (p-1) < 0$.

By Proposition 4, $\|\hat{f}_s - \hat{f}_{s'}\|_{V^\alpha} \leq c_3 \ell^{\kappa_3\varepsilon}|s - s'|$ for $s, s' \in K_\ell$. By construction, $|S_\ell - S_{\ell-1}| \leq \eta_\ell |H(S_{\ell-1}, X_\ell)|$, and assumption (A4) ensures that $|H(S_{\ell-1}, X_\ell)| \leq c\ell^\varepsilon V^\beta(X_\ell)$, so

$$|\hat{f}_{S_\ell}(X_\ell) - \hat{f}_{S_{\ell-1}}(X_\ell)| \leq c_3\ell^{\kappa_3\varepsilon}|S_\ell - S_{\ell-1}|V^\alpha(X_\ell) \leq c_3\ell^{\kappa_3\varepsilon}\eta_\ell c\ell^\varepsilon V^{\alpha+\beta}(X_\ell).$$

Let $k \geq n$. Since $\ell^{(\kappa_3+1)\varepsilon}k^{-\xi} \leq (\ell \vee n)^{(\kappa_3+1)\varepsilon-\xi}$ for $\ell \leq k$, we obtain

$$k^{-\xi}|R_k^{(1)}| \leq k^{-\xi}\sum_{\ell=1}^{k} |\hat{f}_{S_\ell}(X_\ell) - \hat{f}_{S_{\ell-1}}(X_\ell)|$$
$$\leq \tilde{c}\sum_{\ell=1}^{k} (\ell \vee n)^{(\kappa_3+1)\varepsilon-\xi}\eta_\ell V^{\alpha+\beta}(X_\ell)$$

and then by Minkowski's inequality and (11) of Lemma 5,

$$\begin{aligned}
&\mathbb{E}\Big[\max_{n\leq k\leq m} k^{-\xi p}|R_k^{(1)}|^p\Big] \\
&\quad\leq \mathbb{E}\left[\sum_{\ell=1}^{m} \tilde{c}(\ell \vee n)^{(\kappa_3+1)\varepsilon-\xi}\eta_\ell V^{(\alpha+\beta)p}(X_\ell)\right]^p \\
&\quad\leq \tilde{c}\left[\sum_{\ell=1}^{m} \big(\mathbb{E}[(\ell \vee n)^{(\kappa_3+1)\varepsilon-\xi}\eta_\ell V^{\alpha+\beta}(X_\ell)]^p\big)^{1/p}\right]^p \\
&\quad\leq \tilde{c}\left[\sum_{\ell=1}^{\infty} (\ell \vee n)^{(\kappa_3+1+2\alpha+2\beta)\varepsilon-\xi}\eta_\ell\right]^p V^{(\alpha+\beta)p}(x_0).
\end{aligned}\tag{20}$$

Finally, consider $R_k^{(2)}$. From Proposition 4, we have that $\|P_{S_k}\hat{f}_{S_k}(X_k)\|_{V^\alpha} \leq c_3 k^{\kappa_3\varepsilon}$, and by (12) of Lemma 5,

$$\mathbb{E}\Big[\max_{n\leq k\leq m} k^{-\xi p}|P_{S_k}\hat{f}_{S_k}(X_k)|^p\Big] \leq c_3^p\mathbb{E}\Big[\max_{n\leq k\leq m}\big(k^{(\kappa_3\varepsilon-\xi)/\alpha}V(X_k)\big)^{\alpha p}\Big]$$
$$\leq c_3^p c_4^{\alpha p}\left(\sum_{k=n}^{m} k^{(\kappa_3\varepsilon-\xi)/\alpha+2\varepsilon}\right)^{\alpha p} V^{\alpha p}(x_0)$$
$$\leq \tilde{c}n^{(\kappa_3+2\alpha)p\varepsilon+(\alpha-\xi)p}V^{\alpha p}(x_0)$$

since $(\kappa_3 + 2\alpha)\varepsilon - (\xi - \alpha) < 0$. So, we have that

$$\begin{aligned}
\mathbb{E}\Big[\sup_{k\ge n} k^{-\xi p}|R_k^{(2)}|^p\Big] &\le 2^p\mathbb{E}\Big[\sup_{k\ge n} k^{-\xi p}\big(|P_{S_0}\hat{f}_{S_0}(X_0)|^p + |P_{S_k}\hat{f}_{S_k}(X_k)|^p\big)\Big] \\
&\le 2^p\mathbb{E}\Big[|P_{S_0}\hat{f}_{S_0}(X_0)|^p + \sup_{k\ge n} k^{-\xi p}|P_{S_k}\hat{f}_{S_k}(X_k)|^p\Big] \qquad (21)\\
&\le \tilde{c} n^{(\kappa_3+2\alpha)p\varepsilon+(\alpha-\xi)p} V^{\alpha p}(x_0).
\end{aligned}$$

The estimates (16) and (17) follow by Markov's inequality from (20) and (21). □

The proof of Theorem 2 follows as a straightforward application of Proposition 6.

PROOF OF THEOREM 2. Let $\delta > 0$, and denote

$$B_n^{(\delta)} := \bigg\{\omega \in \Omega : \sup_{k\ge n}\frac{1}{k}\bigg|\sum_{j=1}^{k}[f(X_j) - \pi(f)]\bigg| \ge \delta\bigg\}.$$

Since $\|f\|_{V^\alpha} < \infty$ by assumption, we may consider the family $\{f_s\}_{s\in\mathbb{S}}$ with $f_s \equiv f$ for all $s \in \mathbb{S}$. Then, we have by decomposition (14) that

$$\mathbb{P}\big(B_n^{(\delta)}\big) \le \mathbb{P}\bigg[\sup_{k\ge n}\frac{|M_k|}{k} \ge \frac{\delta}{3}\bigg] + \mathbb{P}\bigg[\sup_{k\ge n}\frac{|R_k^{(1)}|}{k} \ge \frac{\delta}{3}\bigg] + \mathbb{P}\bigg[\sup_{k\ge n}\frac{|R_k^{(2)}|}{k} \ge \frac{\delta}{3}\bigg]. \qquad (22)$$

We select $p \in (1, (\alpha+\beta)^{-1})$ so that $\kappa_*\varepsilon < (1 - 1/p)$, and let $\xi = 1$. Then, Proposition 6 readily implies that the first and the third terms in (22) converge to zero as $n \to \infty$. For the second term, consider

$$\sum_{j=1}^{\infty}(j \vee n)^{\kappa_*\varepsilon-1}\eta_j = n^{\kappa_*\varepsilon-1}\sum_{j=1}^{n}\eta_j + \sum_{j=n+1}^{\infty} j^{\kappa_*\varepsilon-1}\eta_j,$$

where the second term converges to zero by assumption, and the first term by Kronecker's lemma. There is an increasing sequence $(n_k)_{k\ge1}$ such that $\mathbb{P}(B_{n_k}^{(1/k)}) \le k^{-2}$. Denoting $B := \bigcap_{m=1}^{\infty}\bigcup_{k=m}^{\infty} B_{n_k}^{(1/k)}$, the Borel–Cantelli lemma implies that $P(B^{\complement}) = 1$, and for all $\omega \in B^{\complement}$, (5) holds. □

4. Bound for the growth rate. In this section, we assume that $\mathbb{X}$ is a normed space, and establish a bound for the growth rate of the chain $(\|X_n\|)_{n\ge1}$, based on a general drift condition. The bound assumes little structure; one must have a drift function V that grows rapidly enough, and that the expected growth of $V(X_n)$ is moderate.

PROPOSITION 7. *Suppose that there is* $V : \mathbb{X} \to [1, \infty)$ *such that the bound*

$$P_s V(x) \le V(x) + b \qquad (23)$$

holds for all $(x, s) \in \mathbb{X} \times \mathbb{S}$, where $b < \infty$ is a constant independent of s. Suppose also that V grows rapidly enough so that

$$\|x\| \geq u \quad \Longrightarrow \quad V(x) \geq r(u) \tag{24}$$

for all $u \geq 0$, where $r : [0, \infty) \to [0, \infty)$ is a function growing faster than any polynomial, that is, for any $p > 0$ there is a $c = c(p) < \infty$ such that

$$\sup_{u \geq 1} \frac{u^p}{r(u)} \leq c. \tag{25}$$

Then, for any $\varepsilon > 0$, there is an a.s. finite $A = A(\omega, \varepsilon)$ such that

$$\|X_n\| \leq A n^{\varepsilon}.$$

PROOF. To start with, (23) implies for $n \geq 1$

$$\begin{aligned}\mathbb{E}[V(X_n)] &= \mathbb{E}[\mathbb{E}[V(X_n)|\mathcal{F}_{n-1}]] = \mathbb{E}[P_{S_{n-1}} V(X_{n-1})] \leq \mathbb{E}[V(X_{n-1})] + b \\ &\leq \cdots \leq V(x_0) + bn \leq \tilde{b} V(x_0) n,\end{aligned}$$

where $\tilde{b} := b + 1$. Now, with fixed $a \geq 1$, we can bound the probability of $\|X_n\|$ ever exceeding an^{ε} as follows

$$\begin{aligned}\mathbb{P}\biggl(\max_{1 \leq n \leq m} \frac{\|X_n\|}{n^{\varepsilon}} \geq a\biggr) &\leq \sum_{n=1}^{m} \mathbb{P}(\|X_n\| \geq an^{\varepsilon}) \leq \sum_{n=1}^{\infty} \mathbb{P}\bigl(V(X_n) \geq r(an^{\varepsilon})\bigr) \\ &\leq \sum_{n=1}^{\infty} \frac{\mathbb{E}[V(X_n)]}{r(an^{\varepsilon})} \leq \tilde{b} V(x_0) \sum_{n=1}^{\infty} \frac{n}{r(an^{\varepsilon})} \\ &\leq \frac{\tilde{b} V(x_0) c}{a^{3/\varepsilon}} \sum_{n=1}^{\infty} n^{-2} \xrightarrow{a \to \infty} 0,\end{aligned}$$

where we use Markov's inequality, and $c = c(3/\varepsilon) < \infty$ is from the application of (25). $\square$

We record the following easy lemma, dealing with a particular choice of $V(x)$, for later use in Section 5.

LEMMA 8. *Assume that the target density π is differentiable, bounded, bounded away from zero on compact sets, and satisfies the following radial decay condition:*

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

Then, for $V(x) = c_V \pi^{-1/2}(x)$, the bound (24) applies with a function $r(u) := ce^{\gamma u}$ for some $\gamma, c > 0$, satisfying (25).

PROOF. Let $R \geq 1$ be such that $\sup_{\|x\| \geq R} \frac{x}{\|x\|} \cdot \nabla \log \pi(x) \leq -\gamma$ for some $\gamma > 0$. Assume $y \in \mathbb{R}^d$ and $\|y\| \geq 2R$, and write $y = (1+a)x$, where $\|x\| = R$ and $a = \frac{\|y\|}{R} - 1 \geq 1$. Denote $h(x) := \log \pi(x)$, and write

$$\log \frac{\pi(y)}{\pi(x)} = \int_1^{1+a} x \cdot \nabla h(tx)\, dt \leq -\gamma a.$$

We have that

$$V(y) = c_V \pi(x)^{-1/2} \left(\frac{\pi(y)}{\pi(x)} \right)^{-1/2} \geq c_V e^{\gamma a/2} \inf_{\|x\|=R} \pi(x)^{-1/2} \geq c e^{\gamma/(4R)\|y\|}$$

and, since π is bounded away from zero on $\{x : \|x\| < 2R\}$, we can select $c > 0$ such that the bound applies to all $y \in \mathbb{R}^d$. $\square$

5. Ergodicity result for adaptive metropolis. We start this section by outlining the original Adaptive Metropolis (AM) algorithm [9]. The AM chain starts from a point $X_0 \equiv x_0 \in \mathbb{R}^d$, and we have an initial covariance $\Sigma_0 \in \mathcal{C}^d$ where $\mathcal{C}^d \subset \mathbb{R}^{d \times d}$ stands for the symmetric and positive definite matrices. We generate, recursively, for $n \geq 0$,

$$X_{n+1} \sim P_{\theta \Sigma_n}(X_n, \cdot), \tag{26}$$

$$\Sigma_{n+1} = \begin{cases} v_0, & 0 \leq n \leq N_b - 1, \\ \mathrm{Cov}(X_0, \dots, X_n) + \kappa I, & n \geq N_b, \end{cases} \tag{27}$$

where $\theta > 0$ is a parameter, $N_b \geq 2$ is the length of the burn-in, $\kappa > 0$ is a small constant, I is an identity matrix and $P_v(x, \cdot)$ is a Metropolis transition probability defined as

$$\begin{aligned} P_v(x, A) := \mathbb{1}_A(x) \left[1 - \int \left(1 \wedge \frac{\pi(y)}{\pi(x)} \right) q_v(y - x)\, dy \right] \\ + \int_A \left(1 \wedge \frac{\pi(y)}{\pi(x)} \right) q_v(y - x)\, dy, \end{aligned} \tag{28}$$

where the proposal density q_v is the Gaussian density with zero mean and covariance $v \in \mathcal{C}^d$.

In this paper, just for notational simplicity (see Remark 9), we consider a slight modification of the AM chain. First, we do not consider a burn-in period, that is, let $N_b = 0$, and let $\Sigma_0 \geq \kappa I$. Instead of (27), we construct Σ_n recursively for $n \geq 1$ as

$$\Sigma_n = \frac{n}{n+1} \Sigma_{n-1} + \frac{1}{n+1} [(X_n - \overline{X}_{n-1})(X_n - \overline{X}_{n-1})^T + \kappa I], \tag{29}$$

where $\overline{X}_n$ denotes the average of $X_0, \dots, X_n$.

REMARK 9. The original AM process uses the unbiased estimate of the covariance matrix. In this case, the recursion formula for Σ_n, when $n \geq N_b + 2$, has the form

$$\Sigma_n = \frac{n-1}{n}\Sigma_{n-1} + \frac{1}{n+1}[(X_n - \overline{X}_{n-1})(X_n - \overline{X}_{n-1})^T + \kappa I]. \tag{30}$$

This recursion can also be formulated in our framework described in Section 2 by simply introducing a sequence of adaptation functions $H_n(s,x)$. Our proof applies with obvious changes. However, in the present paper, we prefer (29) for simpler notation. Also, from a practical point of view, observe that (29) differs from (30) by a factor smaller than $n^{-2}\Sigma_{n-1}$ whence it is mostly a matter of taste whether to use (29) or (30).

In the notation of the general adaptive MCMC framework in Section 2, we have the state space $\mathbb{X} := \mathbb{R}^d$. The adaptation parameter $S_n = (S_n^{(m)}, S_n^{(v)})$ consists of the mean $S_n^{(m)}$ and the covariance $S_n^{(v)}$, having values in $(S_n^{(m)}, S_n^{(v)}) \in \mathbb{S} := \mathbb{R}^d \times \mathcal{C}^d$. The space $\overline{\mathbb{S}} := \mathbb{R}^d \times \mathbb{R}^{d\times d} \supset \mathbb{S}$ is equipped with the norm $|s| := \|s^{(m)}\| \vee \|s^{(v)}\|$ where we use the Euclidean norm, and the matrix norm $\|A\|^2 := \operatorname{trace}(A^T A)$, respectively. The Metropolis kernel P_s is defined as in (28), with the definition $q_s := q_{s^{(v)}}$ for $s \in \mathbb{S}$. The adaptation function H is defined for $s = (s^{(m)}, s^{(v)})$ as

$$H(s,x) := \begin{bmatrix} x - s^{(m)} \\ (x - s^{(m)})(x - s^{(m)})^T - s^{(v)} + \kappa I \end{bmatrix},$$

and the adaptation weights are $\eta_n := (n+1)^{-1}$.

We now formulate our ergodicity result for the AM chain.

THEOREM 10. *Assume π is positive, bounded, bounded from below on compact sets, differentiable and*

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|^\rho} \cdot \nabla \log \pi(x) = -\infty \tag{31}$$

for some $\rho > 1$. Moreover, assume that π has regular contours

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|} \cdot \frac{\nabla\pi(x)}{\|\nabla\pi(x)\|} < 0. \tag{32}$$

Define $V(x) := c_V \pi^{-1/2}(x)$ with $c_V = (\sup_x \pi(x))^{1/2}$. Then, for any f with $\|f\|_{V^\alpha} < \infty$ where $0 \leq \alpha < 1$,

$$\frac{1}{n}\sum_{k=1}^n f(X_k) \xrightarrow{n\to\infty} \pi(f) \tag{33}$$

almost surely.

REMARK 11. If the conditions of Theorem 10 are satisfied, the function $V(x)$ grows faster than an exponential, and hence (33) holds for exponential moments. In particular, (33) holds for power moments, that is, for $f(x) = \|x\|^p$ for any $p \geq 0$, and therefore also $S_n \to (m_\pi, v_\pi + \kappa I)$ where m_π and v_π are the mean and covariance of π.

The proof of Theorem 10 is postponed to the end of this section. We start by a simple lemma bounding the growth rate of the AM chain.

LEMMA 12. *If the conditions of Proposition 7 are satisfied for an AM chain, then for any $\varepsilon > 0$, there is an a.s. finite $A = A(\omega, \varepsilon)$ such that*

$$\|S_n^{(m)}\| \leq An^{\varepsilon}, \qquad \|S_n^{(v)}\| \leq An^{\varepsilon}.$$

PROOF. Since the AM recursion is a convex combination, this is a straightforward corollary of Proposition 7. $\square$

Next, we show that each of the Metropolis kernels used by the AM algorithm satisfy a geometric drift condition, and bound the constants of geometric drift. The result in Proposition 15 is similar to the results obtained in [10, 15], with the exception that we have a common minorization set C for all proposal scalings. We start by two lemmas. We define $\overline{B}(x, r) := \{y \in \mathbb{R}^d : \|x - y\| \leq r\}$.

LEMMA 13. *Assume $E \subset \mathbb{R}^d$ is measurable and $A \subset \mathbb{R}^d$ compact, given as*

$$A := \{ru : u \in S^d, 0 \leq r \leq g(u)\},$$

where $S^d := \{u \in \mathbb{R}^d : \|u\| = 1\}$ is the unit sphere, and $g : S^d \to [b, \infty)$ is a measurable function parameterising the boundary ∂A, with some $b > 0$.

For any $\varepsilon > 0$, define $B_\varepsilon := \{ru : u \in S^d, g(u) < r \leq g(u) + \varepsilon\}$. Then, for all $\tilde{\varepsilon} > 0$, there is a $\tilde{b} = \tilde{b}(\tilde{\varepsilon}) \in (0, \infty)$ such that for all $0 < \varepsilon < \tilde{\varepsilon}$ and for all $\lambda \geq 3\varepsilon$, it holds that

$$|E \cap B_\varepsilon| \leq \big|\big(E \oplus \overline{B}(0, \lambda)\big) \cap A\big|,$$

whenever $b \geq \tilde{b}$. Above, $A \oplus B := \{x + y : x \in A, y \in B\}$ stands for the Minkowski sum.

PROOF. See Figure 1 for an illustration of the situation. Denote by $S^* := \{u \in S^d : \exists r > 0, ur \in E \cap B_\varepsilon\}$ the projection of the set $E \cap B_\varepsilon$ onto S^d. Then we have $E \cap B_\varepsilon \subset \{ru : u \in S^*, g(u) < r \leq g(u) + \varepsilon\}$ and $A \supset \{ru : u \in S^*, 0 \leq r \leq g(u)\}$. Now, for $\varepsilon \leq \lambda \leq g(u)$, we have

$$\big((E \cap B_\varepsilon) \oplus \overline{B}(0, \lambda)\big) \cap A \supset \{ru : u \in S^*, g(u) - \lambda + \varepsilon \leq r \leq g(u)\} =: G,$$

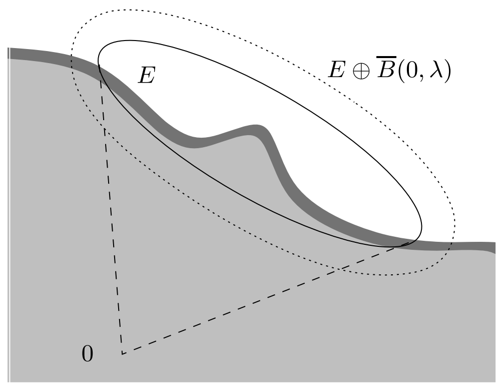

FIG. 1. *Illustration of the boundary estimate. The set A is in light grey, and the set B_ε in dark gray.*

for let $ru \in G$, then there is $g(u) < \tilde{r} \le g(u) + \varepsilon$ such that $\tilde{r}u \in E \cap B_\varepsilon$, and we can write $ru = \tilde{r}u + (r - \tilde{r})u$, where $(r - \tilde{r})u \in \overline{B}(0, \lambda)$. Clearly, $E \oplus \overline{B}(0, \lambda) \supset (E \cap B_\varepsilon) \oplus \overline{B}(0, \lambda)$, and we can estimate

$$\begin{aligned}
&\big|\big(E \oplus \overline{B}(0, \lambda)\big) \cap A\big| - |E \cap B_\varepsilon| \\
&\quad \ge \int_{S^*} \int_{g(u)-2\varepsilon}^{g(u)} r^{d-1}\, dr - \int_{g(u)}^{g(u)+\varepsilon} r^{d-1}\, dr\, \mathcal{H}^{d-1}(du) \\
&\quad = \frac{1}{d} \int_{S^*} 2(g(u))^d - \big(g(u) - 2\varepsilon\big)^d - \big(g(u) + \varepsilon\big)^d \mathcal{H}^{d-1}(du),
\end{aligned}$$

where $\mathcal{H}^{d-1}$ stands for the $(d-1)$-dimensional Hausdorff measure. This integral is nonnegative for all $0 \le \varepsilon \le c_d b$, for some constant c_d depending only on the dimension d, namely let $h(\varepsilon) := (y - 2\varepsilon)^d + (y + \varepsilon)^d$. The mean value theorem implies that for some $0 \le \varepsilon' \le \varepsilon$, one has

$$h(0) - h(\varepsilon) = \varepsilon d (y - 2\varepsilon')^{d-1} \left[2 - \left(\frac{y + \varepsilon'}{y - 2\varepsilon'} \right)^{d-1} \right] \ge 0,$$

whenever $\varepsilon \le c_d y$. $\square$

LEMMA 14. *Let $f(x) := x e^{-x^2/2}$. For any $0 < \varepsilon < 1/8$, the following estimates hold*:

$$2f(x + \varepsilon) - f(x) \ge \frac{x}{8} \qquad \textit{for all } 0 < x \le \frac{1}{2}$$

and

$$\int_0^\infty \big([2f(x + \varepsilon) - f(x)] \wedge 0\big)\, dx \ge -e^{-c\varepsilon^{-2}}$$

for some constant $c > 0$.

PROOF. We can write

$$2f(x+\varepsilon) - f(x) = e^{-x^2/2}[2(x+\varepsilon)e^{-x\varepsilon-\varepsilon^2/2} - x],$$

which is positive whenever $e^{-x\varepsilon-\varepsilon^2/2} \geq 2/3$, holding at least for all $0 \leq x \leq x^*$, with

$$x^* = \frac{\log(3/2)}{\varepsilon} - \frac{\varepsilon}{2} \geq \frac{1}{4\varepsilon}.$$

Now, $x^* \geq 1/2$ and we can estimate

$$2f(x+\varepsilon) - f(x) \geq \frac{1}{4}xe^{-x^2/2} \geq \frac{x}{8}$$

for all $0 < x \leq 1/2$. Also,

$$\int_0^\infty \big([2f(x+\varepsilon) - f(x)] \wedge 0\big)\, dx \geq -\int_{x^*}^\infty xe^{-x^2/2}\, dx = -e^{-c\varepsilon^{-2}}$$

with $c = 1/32$. $\square$

PROPOSITION 15. *Assume that π satisfies the conditions in Theorem 10 and $\kappa > 0$. 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 the Metropolis transition probability P_v in (28) and for all $v \in \mathcal{C}^d$ with all eigenvalues greater than $\kappa > 0$, it holds that*

$$P_v V(x) \leq \lambda_v V(x) + b\mathbb{1}_C(x) \qquad \forall x \in \mathbb{X}, \tag{34}$$

$$P_v(x, B) \geq \delta_v \nu(B) \qquad \forall x \in C, \forall B \subset \mathbb{X}, \tag{35}$$

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

$$(1-\lambda_v)^{-1} \vee \delta_v^{-1} \leq c \det(v)^{1/2}$$

for some constant $c \geq 1$.

PROOF. Define the sets $A_x := \{y : \pi(y) \geq \pi(x)\}$ and its complement $R_x := \{y : \pi(y) < \pi(x)\}$, which are the regions of almost sure acceptance and possible rejection at x, respectively. Let $R > 1$ be sufficiently large to ensure that for all $\|x\| \geq R$, it holds that

$$\sup_{\|x\|\geq R} \frac{x}{\|x\|} \cdot \frac{\nabla\pi(x)}{\|\nabla\pi(x)\|} < -\gamma \quad \text{and} \quad \sup_{\|x\|\geq R} \frac{x}{\|x\|} \cdot \nabla \log \pi(x) < -\|x\|^{\rho-1}$$

for some $\gamma > 0$. Suppose that the dimension $d \geq 2$. Lemma 22 in Appendix C implies that for R sufficiently large, we have $\overline{B}(0, M^{-1}\|x\|) \subset A_x \subset \overline{B}(0, M\|x\|)$ for all $\|x\| \geq R$ with some constant $M \geq 1$. Moreover, we can parameterize $A_x = \{ru : u \in S^d, 0 \leq r \leq g(u)\}$ where $S^d := \{u \in \mathbb{R}^d : \|u\| = 1\}$ is the unit sphere, and $g : S^d \to [M^{-1}\|x\|, M\|x\|]$.

Consider (34). We may compute

$$\tau_v := 1 - \frac{P_v V(x)}{V(x)}$$
$$= \int_{A_x} \left(1 - \sqrt{\frac{\pi(x)}{\pi(y)}}\right) q_v(y-x)\,dy \tag{36}$$
$$- \int_{R_x} \sqrt{\frac{\pi(y)}{\pi(x)}} \left(1 - \sqrt{\frac{\pi(y)}{\pi(x)}}\right) q_v(y-x)\,dy.$$

In what follows, unless explicitly stated, we assume $\|x\| \geq M(R+1)$. Denote $\varepsilon_x := \|x\|^{-\alpha} < 1$, where $\alpha = (\rho - 1)/2 > 0$. Define $\tilde{A}_x := \{ru : u \in S^d, 0 \leq r \leq g(u) - \varepsilon_x\} \subset A_x$ and $\tilde{R}_x := \{ru : u \in S^d, r \geq g(u) + \varepsilon_x\} \subset R_x$. From (36), we can estimate

$$\tau_v \geq \int \left[\left(1 - \sqrt{\frac{\pi(x)}{\pi(y)}}\right) \mathbb{1}_{\tilde{A}_x}(y) - \frac{1}{4} \mathbb{1}_{R_x \setminus \tilde{R}_x}(y)\right] q_v(y-x)\,dy \tag{37}$$
$$- \sup_{z \in \mathbb{R}^d} q_v(z-x) \int_{\tilde{R}_x} \sqrt{\frac{\pi(y)}{\pi(x)}}\,dy.$$

We estimate the two terms in the right-hand side separately, starting from the first.

Let $h(x) := \log \pi(x)$. Suppose $z \in \tilde{A}_x$, and write $z = (1 - a/\|y\|)y$ for some $y \in \partial A_x$ and $\varepsilon_x \leq a \leq \|y\|$. Assume for a moment $\|z\| \geq R$. Then, h is decreasing on the line segment from z to y, and we can estimate

$$\frac{\pi(x)}{\pi(z)} = \frac{\pi(y)}{\pi(z)} = e^{h(y)-h(z)} = e^{\int_{\|y\|-a}^{\|y\|} y/\|y\| \cdot \nabla h(ty/\|y\|)\,dt} \leq e^{\int_{\|y\|-\varepsilon_x}^{\|y\|} y/\|y\| \cdot \nabla h(ty/\|y\|)\,dt}$$
$$\leq e^{-\varepsilon_x (\|y\| - \varepsilon_x)^{\rho-1}} \leq e^{-\varepsilon_x \|x\|^{\rho-1}/(2M)^{\rho-1}} = e^{-\|x\|^{\alpha}/(2M)^{\rho-1}}.$$

Hence, in this case, $\pi(x)/\pi(z) \leq 1/4$ assuming $\|x\| \geq R_2$ for sufficiently large $R_2 \geq R$. If $\|z\| < R$, then there is z' such that $\|z'\| = R$ and the estimate above holds for z'. Consequently,

$$\frac{\pi(x)}{\pi(z)} = \frac{\pi(y)}{\pi(z')} \frac{\pi(z')}{\pi(z)} \leq e^{-\|x\|^{\alpha}/(2M)^{\rho-1}} \frac{\sup_{\|w\| \leq R} \pi(w)}{\inf_{\|w\| \leq R} \pi(w)} \leq \frac{1}{4}, \tag{38}$$

whenever $\|x\| \geq R_2$ by increasing R_2 if needed. In conclusion, we have shown that for $\|x\| \geq R_2$, it holds that $(1 - \sqrt{\pi(x)/\pi(y)}) \geq 1/2$ for all $y \in \tilde{A}_x$.

By Fubini's theorem, we can write for positive f that

$$\int f(z+x) q_v(z)\,dx = \frac{c_d}{\sqrt{\det(v)}} \int_0^1 \int_{\{e^{-1/2 z^T v^{-1} z} \geq t\}} f(z+x)\,dz\,dt$$
$$= \frac{c_d}{\sqrt{\det(v)}} \int_0^\infty \int_{E_u} f(y)\,dy\, u e^{-u^2/2}\,du,$$

where $c_d = (2\pi)^{-d/2}$ and $E_u := \{z + x : z^T v^{-1} z \le u^2\}$. Consequently, for $\|x\| \ge R_2$, we can estimate the first term of (37) from below by

$$\begin{aligned}
&\int_0^\infty \left(\frac{|E_u \cap \tilde{A}_x|}{2} - \frac{|E_u \cap (R_x \setminus \tilde{R}_x)|}{4}\right) u e^{-u^2/2}\, du \\
&\quad \ge \frac{1}{4}\int_0^\infty 2|E_{u+a} \cap \tilde{A}_x|(u+a)e^{-(u+a)^2/2} - |E_u \cap (R_x \setminus \tilde{R}_x)| u e^{-u^2/2}\, du \\
&\quad \ge \frac{1}{4}\int_0^\infty 2\big|\big(E_u \oplus \overline{B}(0, \kappa^{1/2}a)\big) \cap \tilde{A}_x\big|(u+a)e^{-(u+a)^2/2} \\
&\qquad\qquad - |E_u \cap B_\varepsilon| u e^{-u^2/2}\, du
\end{aligned}$$

for any $a \ge 0$, since simple computation shows that $E_u \oplus \overline{B}(0, \kappa^{1/2}a) = \{x + y : x \in E_u, y \in \overline{B}(0, \kappa^{1/2}a)\} \subset E_{u+a}$, and as we may write $\tilde{A}_x = \{ru : u \in S^d, 0 \le r \le \tilde{g}(u)\}$ where $\tilde{g}(u) = g(u) - \varepsilon_x$, we obtain that $R_x \setminus \tilde{R}_x \subset \{ru : u \in S^d, \tilde{g}(u) \le r \le \tilde{g}(u) + 2\varepsilon_x\} =: B_\varepsilon$. We set $a = 6\kappa^{-1/2}\varepsilon_x$ and apply Lemma 13 with the choice $\varepsilon = 2\varepsilon_x$ and $\lambda = 6\varepsilon_x$,

$$\begin{aligned}
&\int_0^\infty \left(\frac{|E_u \cap \tilde{A}_x|}{2} - \frac{|E_u \cap (R_x \setminus \tilde{R}_x)|}{4}\right) u e^{-u^2/2}\, du \\
&\quad \ge \frac{1}{4}\int_0^\infty |[E_u \oplus \overline{B}(0, 6\varepsilon_x)] \cap \tilde{A}_x|\big[2(u+a)e^{-(u+a)^2/2} - u e^{-u^2/2}\big]\, du \\
&\quad \ge \frac{1}{4}\int_{1/4}^{1/2} |E_u \cap \tilde{A}_x| \frac{u}{8}\, du - |\tilde{A}_x| e^{-c_1 \varepsilon_x^{-2}} \\
&\quad \ge c_2 |E_{1/4} \cap \tilde{A}_x| - M^d \|x\|^d e^{-c_1 \|x\|^\alpha}
\end{aligned}$$

by Lemma 14, for sufficiently large $\|x\|$, and since E_u are increasing with respect to u. We have that $E_{1/4} \supset \overline{B}(x, \kappa^{1/2}/4)$. If $\|x\| \to \infty$, then $\varepsilon_x \to 0$ and also $|\overline{B}(x, \kappa^{1/2}/4) \cap \tilde{A}_x| - |\overline{B}(x, \kappa^{1/2}/4) \cap A_x| \to 0$. Moreover, it holds that $|\overline{B}(x, \kappa^{1/2}/4) \cap A_x| \ge c_3 > 0$ (see the proof of Theorem 4.3 in [10]). So, for large enough $\|x\|$, there is a $c_4 > 0$ so that $|E_{1/4} \cap \tilde{A}_x| \ge c_4$. To sum up, by choosing R_3 to be sufficiently large, we obtain that the first part of (37) is at least $c_5(\det(v))^{-1/2}$ for all $\|x\| \ge R_3$, with a $c_5 > 0$.

Next, we turn to the second term of (37). We obtain by polar integration that

$$\begin{aligned}
\int_{\tilde{R}_x} \sqrt{\frac{\pi(y)}{\pi(x)}}\, dy &= \int_{S^d} \int_{g(u)+\varepsilon_x}^\infty r^{d-1} e^{1/2h(ru) - 1/2h(g(u)u)}\, dr\, \mathcal{H}^{d-1}(du) \\
&\le c_d' \sup_{M^{-1}\|x\| \le w \le M\|x\|} \int_{w+\varepsilon_x}^\infty r^{d-1} e^{-1/2 \int_w^r t^{\rho-1}\, dt}\, dr,
\end{aligned}$$

where $\mathcal{H}^{d-1}$ is the $(d-1)$-dimensional Hausdorff measure, and $c_d' = \mathcal{H}^{d-1}(S^d)$. Denote $T(w, r) := r^{d-1} e^{-1/4 \int_w^r t^{\rho-1}\, dt}$ and let us estimate the latter integral from above by

$$\int_{w+\varepsilon_x}^{\infty} e^{-1/4\int_w^r t^{\rho-1}\,dt}\,dr \sup_{r\ge w+\varepsilon_x} T(w,r) \le \int_w^{\infty} e^{-w^{\rho-1}/4(r-w)}\,dr \sup_{r\ge w+\varepsilon_x} T(w,r)$$
$$\le 4M^{\rho-1}\|x\|^{1-\rho} \sup_{r\ge w+\varepsilon_x} T(w,r)$$

for any $w \ge M^{-1}\|x\|$. Suppose first $w+\varepsilon_x \le r \le 2w$, then

$$T(w,r) \le (2w)^{d-1}e^{-1/4\varepsilon_x w^{\rho-1}} \le (2M)^{d-1}\|x\|^{d-1}e^{-1/4M^{1-\rho}\|x\|^{\alpha}} \le c_6$$

for any $M^{-1}\|x\| \le w \le M\|x\|$. For any $r > 2w$ and $w \ge 1$, we have

$$T(w,r) \le r^{d-1}e^{-1/4r/2w^{\rho-1}} \le r^{d-1}e^{-r/8} \le c_7.$$

Put together, letting $R_4 \ge R_3$ to be sufficiently large, we obtain that $\tau_v \ge c_8(\det(v))^{-1/2}$ with $c_8 = c_5/2$ for all $\|x\| \ge R_4$.

To sum up, by setting $C = \overline{B}(0, R_4)$, we get that for all $v \in \mathcal{C}^d$ with eigenvalues bounded from below by κ, the estimate $P_v V(x) \le \lambda_v V(x)$ holds for $x \notin C$ with $\lambda_v := 1 - c_8 \det(v)^{-1/2}$ satisfying $(1-\lambda_v)^{-1} \le c_8^{-1}\det(v)^{1/2}$. For $x \in C$, we have by (36) that $P_v V(x) \le 2V(x) \le 2\sup_{z\in C} V(z) \le b < \infty$, so (34) holds. In the one-dimensional case, the above estimates can be applied separately for the tails of the distribution.

Finally, set $\nu(B) := |C|^{-1}|B\cap C|$, and consider the minorization condition (35) for $x \in C$,

$$\begin{aligned}
P_v(x,B) &\ge \int_{B\cap C}\left(1 \wedge \frac{\pi(y)}{\pi(x)}\right) q_v(y-x)\,dy \\
&\ge \frac{c_d}{\sqrt{\det(v)}}\int_{B\cap C}\left(1 \wedge \frac{\pi(y)}{\pi(x)}\right) \inf_{x,y\in C} e^{-1/2(x-y)v^{-1}(x-y)}\,dy \\
&\ge \frac{c_d}{\sqrt{\det(v)}} e^{-1/(2\kappa')\operatorname{diam}(C)^2} \frac{\inf_{z\in C}\pi(z)}{\sup_z \pi(z)}\int_{B\cap C} dy.
\end{aligned}$$

So (35) holds with $\delta_v := c_9 \det(v)^{-1/2}$ for some $c_9 > 0$. Finally, the claim holds with $c := c_8^{-1} \vee c_9^{-1}$. $\square$

Finally, we are ready to prove the strong law of large numbers for the AM process.

PROOF OF THEOREM 10. We start by verifying the strong law of large numbers (33). Fix $t \ge 1$ and consider first the constrained process $(X_n^{(t)}, S_n^{(t)})_{n\ge 0}$ which is defined as the AM chain, but with the constraint sets $K_n^{(t)}$ defined as $K_n^{(t)} := \{s \in \mathbb{S} : |s| \le tn^{\varepsilon'}\}$, with $\varepsilon' = \varepsilon/(2d)$, and $\varepsilon \in (0, \kappa_*^{-1}[(1/2)\wedge(1-\alpha)])$, where κ_* is the independent constant of Theorem 2.

We check that assumptions (A1)–(A4) are satisfied by the constrained process $(X_n^{(t)}, S_n^{(t)})_{n\geq 0}$ for all $t \geq 1$. Condition (A1) is satisfied by construction of the Metropolis kernels P_s. Since $\det(v) \leq \|v\|^d$, Proposition 15 ensures that there is a compact $C \subset \mathbb{R}^d$ such that (A2) holds. For (A3), we refer to [1], Lemma 13, stating that $\|P_s f - P_{s'} f\|_{V^r} \leq 2d\kappa^{-1}\|f\|_{V^r}|s^{(v)} - s'^{(v)}|$ for all $s^{(v)}, s'^{(v)} \in \mathcal{C}^d$ with eigenvalues bounded from below by κ.

Finally, we check that (A4) holds for any $\beta \in (0, 1/2]$. Similarly to [2], we have that

$$
\begin{aligned}
\sup_{s\in K_n^{(t)}} &\|H(s,x)\|_{V^\beta} \\
&= \sup_{s\in K_n^{(t)}} \sup_{x\in\mathbb{R}^d} \frac{|H(s,x)|}{V^\beta(x)} \\
&\leq \|\kappa I\| + \sup_{x\in\mathbb{R}^d} \sup_{s\in K_n^{(t)}} \frac{\|x\| + \|s^{(m)}\| + \|s^{(v)}\| + \|(x - s^{(m)})(x - s^{(m)})^T\|}{V^\beta(x)} \\
&\leq \sqrt{d}\kappa + \sup_{x\in\mathbb{R}^d} \frac{\|x\| + \|x\|^2 + t^2 n^{2\varepsilon'} + 2tn^{\varepsilon'} + 2\|x\|tn^{\varepsilon'}}{V^\beta(x)} \\
&\leq \sqrt{d}\kappa + 7t^2 n^{2\varepsilon'} \sup_{x\in\mathbb{R}^d} \frac{\|x\|^2 \vee 1}{V^\beta(x)} \leq \tilde{c} n^\varepsilon
\end{aligned}
$$

for any $\beta \in (0, 1/2]$ by Lemma 8, where $\tilde{c} = \tilde{c}(t, \beta)$. So, assumption (A4) holds for any $\beta \in (0, 1-\alpha)$. In particular, we can select β so that $\varepsilon < \kappa_*^{-1}[(1/2) \wedge (1 - \alpha - \beta)]$. Clearly, $\sum_k k^{\kappa_*\varepsilon - 1}\eta_k < \sum_k k^{\kappa_*\varepsilon-2} < \infty$, so all the conditions of Theorem 2 are satisfied, implying that the strong law of large numbers holds for the constrained process $(X_n^{(t)}, S_n^{(t)})$ for all $t \geq 1$.

Define $B^{(t)} := \{\forall n \geq 0 : S_n \in K_n^{(t)}\}$. We can construct the constrained processes so that they coincide with the original process in $B^{(t)}$. That is, for $\omega \in B^{(t)}$ we have $(X_n(\omega), S_n(\omega)) = (X_n^{(t)}(\omega), S_n^{(t)}(\omega))$ for all $n \geq 0$. Lemma 12 ensures that we have $\mathbb{P}(\forall n \geq 0 : S_n \in K_n^{(t)}) \geq g(t)$ where $g(t) \to 1$ as $t \to \infty$. As in the proof of Theorem 2, we can use the Borel–Cantelli lemma to deduce that (33) holds almost surely. □

REMARK 16. Since $\varepsilon > 0$ can be selected arbitrarily small in the proof of Theorem 10, it is only required for (33) to hold that the adaptation weights $\eta_n \in (0, 1)$ are decreasing and that $\sum_k k^{\tilde{\varepsilon}-1}\eta_k < \infty$ holds for some $\tilde{\varepsilon} > 0$. In particular, one can choose $\eta_n := (n+1)^{-\gamma}$ for any $\gamma > 0$.

REMARK 17. Condition (31) implies the super-exponential decay of the tails of π:

$$
\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|} \cdot \nabla \log \pi(x) = -\infty. \tag{39}
$$

This condition, with the contour regularity condition (32), are common conditions to ensure geometric ergodicity of a random-walk Metropolis algorithm, and many standard distributions fulfil them [10]. The decay condition (31) is only slightly more stringent than (39).

Finally, we formulate a central limit theorem for the AM algorithm.

THEOREM 18. *Assume π satisfies the conditions of Theorem 10. For any f with $\|f\|_{V^\alpha} < \infty$ for some $0 \le \alpha < 1/2$, where $V(x) := c_V \pi^{-1/2}(x)$ and $c_V = (\sup_x \pi(x))^{1/2}$, it holds that*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} [f(X_k) - \pi(f)] \xrightarrow{n\to\infty} N(0, \sigma^2)$$

in distribution, where $\sigma^2 \in [0, \infty)$ is a constant.

The proof of Theorem 18 follows by the techniques of the present paper applied to [1], Theorem 9. A fully detailed proof can be found in the preprint [16].

APPENDIX A: PROOF OF LEMMA 3

We provide a restatement of a part of a theorem by Meyn and Tweedie [12] before proving Lemma 3. For a more recent work on quantitative convergence bounds, we refer to [6].

THEOREM 19. *Suppose that the following drift and minorization conditions hold:*

$$PV(x) \le \lambda V(x) + b \mathbb{1}_C(x) \qquad \forall x \in \mathbb{X},$$
$$P(x, A) \ge \delta \nu(A) \qquad \forall x \in C, \forall A \subset \mathbb{X}$$

for constants $\lambda < 1$, $b < \infty$ and $\delta > 0$, a set $C \subset \mathbb{X}$ and a probability measure ν on C. Moreover, suppose that $\sup_{x \in C} V(x) \le b$. Then, for all $k \ge 1$,

$$\|P_s^k(x, \cdot) - \pi(\cdot)\|_V \le V(x)(1+\gamma)\frac{\rho}{\rho - \vartheta}\rho^k$$

for any $\rho > \vartheta = 1 - \tilde{M}^{-1}$, for

$$\tilde{M} = \frac{1}{(1-\check{\lambda})^2}[1 - \check{\lambda} + \check{b} + \check{b}^2 + \bar{\zeta}(\check{b}(1-\check{\lambda})\check{b}^2)]$$

defined in terms of

$$\gamma = \delta^{-2}[4b + 2\delta\lambda b],$$
$$\check{\lambda} = (\lambda + \gamma)/(1+\gamma) < 1,$$
$$\check{b} = b + \gamma < \infty$$

and the bound

$$\bar{\zeta} \le \frac{4-\delta^2}{\delta^5}\left(\frac{b}{1-\lambda}\right)^2.$$

PROOF. See [12], Theorem 2.3. $\square$

PROOF OF LEMMA 3. Observe that $P_s V(x) = \mathbb{E}[V(X_{n+1})|X_n = x, S_n = s]$, and therefore by Jensen's inequality, (A2) implies for $x \notin C_n$ that

$$P_s V^r(x) \le (P_s V(x))^r \le \lambda_n^r V^r(x).$$

We can bound $\tilde{\lambda}_n := \lambda_n^r \le (1 - c^{-1}n^{-\varepsilon})^r \le 1 - rc^{-1}n^{-\varepsilon}$ implying

$$(1-\tilde{\lambda}_n)^{-1} \le r^{-1}cn^{\varepsilon},$$

whenever $r \in (0, 1]$. Similarly, for $x \in C_n$, one has $P_s V^r(x) \le (\sup_{z\in C_n} V(z) + b_n)^r \le (2b_n)^r$, so by letting $\tilde{b}_n := (2b_n)^r$, we obtain the drift inequality

$$P_s V^r(x) \le \tilde{\lambda}_n V^r(x) + \tilde{b}_n \mathbb{1}_{C_n}(x),$$

and we can bound $\tilde{b}_n \le (2cn^{\varepsilon})^r$. We have the bound $(1-\tilde{\lambda}_n)^{-1} \vee \tilde{b}_n \le \tilde{c}n^{\varepsilon}$ with some $\tilde{c} = \tilde{c}(c, r) \ge 1$.

Now, we can apply Theorem 19, where we can estimate the constants

$$\begin{aligned}
\gamma_n &= \delta_n^{-2}[4\tilde{b}_n + 2\delta_n\tilde{\lambda}_n\tilde{b}_n] \le (cn^{\varepsilon})^2 6(\tilde{c}n^{\varepsilon}) = a_1 n^{3\varepsilon},\\
\check{b}_n &= \tilde{b}_n + \gamma_n \le (\tilde{c} + a_1)n^{3\varepsilon} \le a_2 n^{3\varepsilon}
\end{aligned}$$

and consequently

$$1 - \check{\lambda}_n = \frac{1-\tilde{\lambda}_n}{1+\gamma_n} \ge \frac{\tilde{c}^{-1}n^{-\varepsilon}}{1+a_1 n^{3\varepsilon}} \ge \frac{\tilde{c}^{-1}}{1+a_1} n^{-4\varepsilon} = a_3^{-1} n^{-4\varepsilon}.$$

Moreover,

$$\bar{\zeta}_n \le \frac{4-\delta_n^2}{\delta_n^5}\left(\frac{\tilde{b}_n}{1-\tilde{\lambda}_n}\right)^2 \le 4(cn^{\varepsilon})^5(\tilde{c}n^{\varepsilon})^2(\tilde{c}n^{\varepsilon})^2 = a_4 n^{9\varepsilon},$$

and then

$$\begin{aligned}
\tilde{M}_n &= \frac{1}{(1-\check{\lambda})^2}\left[1 - \check{\lambda}_n + \check{b}_n + \check{b}_n^2 + \bar{\zeta}_n\big(\check{b}_n(1-\check{\lambda}_n) + \check{b}_n^2\big)\right]\\
&\le (a_3 n^{4\varepsilon})^2[1 + \check{b}_n + \check{b}_n^2 + \bar{\zeta}_n(\check{b}_n + \check{b}_n^2)]\\
&\le (a_3 n^{4\varepsilon})^2(5\bar{\zeta}_n\check{b}_n^2) \le 5a_3^2 n^{8\varepsilon} a_4 n^{9\varepsilon} a_2^2 n^{6\varepsilon} = a_5 n^{23\varepsilon}
\end{aligned}$$

since we can assume that $\check{b}_n, \bar{\zeta}_n \ge 1$. Now,

$$1 - \vartheta_n = \tilde{M}_n^{-1} \ge a_5^{-1} n^{-23\varepsilon}$$

and we can choose $\rho_n \in (\vartheta_n, 1)$ by letting $\rho_n := \frac{1+\vartheta_n}{2}$. We have

$$\rho_n - \vartheta_n = 1 - \rho_n = \tfrac{1}{2}(1-\vartheta_n) \geq \tfrac{1}{2} c_9^{-1} n^{-23\varepsilon} = (a_6 n^{23\varepsilon})^{-1}.$$

Finally, from Theorem 19, one obtains the bound

$$\|P_s^k(x,\cdot) - \pi(\cdot)\|_{V^r} \leq V^r(x) L_n \rho_n^k,$$

where

$$\begin{aligned}
(1-\rho_n)^{-1} &\leq a_6 n^{23\varepsilon}, \\
L_n &= (1+\gamma_n)\frac{\rho_n}{\rho_n - \vartheta_n} \leq (1 + a_1 n^{3\varepsilon})(a_6 n^{23\varepsilon}) \leq a_7 n^{26\varepsilon}
\end{aligned}$$

with $a_7 = (1+a_1)a_6$. This concludes the proof with $\kappa_2 = 26$ and $c_2 = a_7$. □

APPENDIX B: BIRNBAUM AND MARSHALL'S INEQUALITY

THEOREM 20 (Birnbaum and Marshall). *Let $(X_k)_{k=1}^n$ be random variables, such that*

$$\mathbb{E}[|X_k||\mathcal{F}_{k-1}] \geq \psi_k |X_{k-1}|,$$

where $\mathcal{F}_k := \sigma(X_1, \dots, X_k)$, and $\psi_k \geq 0$. Let $a_k > 0$, and define

$$b_k := \max\left\{a_k, a_{k+1}\psi_{k+1}, \dots, a_n \prod_{j=k+1}^{n} \psi_j\right\}$$

for $1 \leq k \leq n$, and $b_{n+1} := 0$. If $p \geq 1$ is such that $\mathbb{E}|X_k|^p < \infty$ for all $1 \leq k \leq n$, then

$$\mathbb{P}\Big(\max_{1\leq k\leq n} a_k |X_k| \geq 1\Big) \leq \sum_{k=1}^{n} (b_k^p - \psi_{k+1}^p b_{k+1}^p)\mathbb{E}|X_k|^p.$$

PROOF. See [7], Theorem 2.1. □

COROLLARY 21. *Let $(M_k)_{k=1}^n$ be a martingale with respect to $(\mathcal{F}_k)_{k=1}^n$. Let $(a_k)_{k=1}^n$ be a strictly positive nonincreasing sequence. If $p \geq 1$ is such that $\mathbb{E}|M_k|^p < \infty$ for all $1 \leq k \leq n$, then for $1 \leq m \leq n$,*

$$\mathbb{P}\Big(\max_{m\leq k\leq n} a_k |M_k| \geq 1\Big) \leq a_n^p \mathbb{E}|M_n|^p + \sum_{k=m}^{n-1} (a_k^p - a_{k+1}^p)\mathbb{E}|M_k|^p.$$

PROOF. By Jensen's inequality,

$$\mathbb{E}[|M_k||\mathcal{F}_{k-1}] \geq |\mathbb{E}[M_k|\mathcal{F}_{k-1}]| = |M_{k-1}|.$$

Define $\psi_k := 1$ for $1 \leq k \leq n$, and $\tilde{a}_k := a_m$ for $1 \leq k \leq m$ and $\tilde{a}_k := a_k$ for $m < k \leq n$. The result follows from Theorem 20. □

APPENDIX C: CONTOUR SURFACE CONTAINMENT

LEMMA 22. *Suppose $A \subset \mathbb{R}^d$ is a smooth surface parameterized by the unit sphere $\mathcal{S}^d$, that is, $A = \{ug(u) : u \in \mathcal{S}^d\}$ with a continuously differentiable radial function $g : \mathcal{S}^d \to (0, \infty)$. Assume also that outer-pointing normal n of A satisfies $n(x) \cdot x/\|x\| \geq \beta$ for all $x \in A$ with some constant $\beta > 0$. There is a constant $M < \infty$ depending only on β such that for any $x, y \in A$, it holds that $M^{-1} \leq \|x\|/\|y\| \leq M$.*

PROOF. Consider first the two-dimensional case. Let x and y be two distinct points in A. We employ polar coordinates, thus let $u(\theta)r(\theta) \in A$ with $u(\theta) := [\cos(\theta), \sin(\theta)]^T$ and $r(\theta) := g(u(\theta))$ so that $u(\theta_1)r(\theta_1) = x$ and $u(\theta_2)r(\theta_2) = y$ with $\theta_1, \theta_2 \in [0, 2\pi)$.

Let $\alpha(\theta)$ stand for the (smaller) angle between $u(\theta)$ and the normal of the curve A, that is, the curve parametrized by $\theta \to u(\theta)r(\theta)$. Our assumption says that $|\alpha(t)| \leq \alpha_0 := \arccos(\beta) < \pi/2$ for all $\theta \in [0, 2\pi]$. On the other hand, an elementary computation shows that

$$\tan(\alpha(\theta)) = \frac{r'(\theta)}{r(\theta)},$$

and hence we have $|\frac{d}{d\theta} \log r(\theta))| = |r'(\theta)/r(\theta)| \leq \tan \alpha_0$ uniformly. We may estimate $|\log\|x\| - \log\|y\|| \leq 2\pi \tan(\alpha_0)$ yielding the claim with $M = e^{2\pi \tan \alpha_0}$.

For $d \geq 3$, take the plane T containing the origin and the points x and y. This reduces the situation to two dimensions, since $A \cap T$ inherits the given normal condition of the surface and the radius vector. □

Acknowledgments. We thank the anonymous referees for a careful review and comments improving the paper significantly. We also thank Gersende Fort for useful comments.

REFERENCES

[1] ANDRIEU, C. and MOULINES, É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.* **16** 1462–1505. MR2260070

[2] ANDRIEU, C., MOULINES, É. and PRIOURET, P. (2005). Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.* **44** 283–312 (electronic). MR2177157

[3] ANDRIEU, C. AND ROBERT, C. P. (2001). Controlled MCMC for optimal sampling. Technical Report Ceremade 0125, Univ. Paris Dauphine.

[4] ANDRIEU, C. and THOMS, J. (2008). A tutorial on adaptive MCMC. *Statist. Comput.* **18** 343–373.

[5] ATCHADÉ, Y. F. and ROSENTHAL, J. S. (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli* **11** 815–828. MR2172842

[6] BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987

[7] BIRNBAUM, Z. W. and MARSHALL, A. W. (1961). Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Statist.* **32** 687–703. MR0148106

[8] HAARIO, H., LAINE, M., LEHTINEN, M., SAKSMAN, E. and TAMMINEN, J. (2004). Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 591–607. MR2088292

[9] HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An adaptive metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504

[10] JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030

[11] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21** 1087–1092.

[12] MEYN, S. P. and TWEEDIE, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* **4** 981–1011. MR1304770

[13] ROBERTS, G. O. and ROSENTHAL, J. S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.* **44** 458–475. MR2340211

[14] ROBERTS, G. O. and ROSENTHAL, J. S. (2009). Examples of adaptive MCMC. *J. Comput. Graph. Statist.* **18** 349–367.

[15] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158

[16] SAKSMAN, E. and VIHOLA, M. (2010). On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. Preprint. Available at arXiv:0806.2933v4.

DEPARTMENT OF MATHEMATICS AND STATISTICS
P.O. BOX 68
FI-00014 UNIVERSITY OF HELSINKI
FINLAND
E-MAIL: eero.saksman@helsinki.fi

DEPARTMENT OF MATHEMATICS AND STATISTICS
P.O. BOX 35 (MAD)
FI-40014 UNIVERSITY OF JYVÄSKYLÄ
FINLAND
E-MAIL: matti.vihola@iki.fi
URL: http://iki.fi/mvihola