ON THE ERGODICITY OF THE ADAPTIVE METROPOLIS ALGORITHM ON UNBOUNDED DOMAINS

EERO SAKSMAN AND MATTI VIHOLA

Abstract. This paper describes sufficient conditions to ensure the correct ergodicity of the Adaptive Metropolis (AM) algorithm of Haario, Saksman, and Tamminen [9], for target distributions with a non-compact support. The conditions ensuring a strong law of large numbers and a central limit theorem 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, and 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 [1].

1. Introduction

The Markov chain Monte Carlo (MCMC) method, first proposed by [13], is a commonly used device for numerical approximation of integrals of the type

$$\pi(f) = \int f(x)\pi(x)\mathrm{d}x$$

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 realisation 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, e.g. a zero-mean Gaussian, and
(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 $X_n = X_{n-1}$.

Date: June 15, 2010.

2000 *Mathematics Subject Classification.* Primary 65C05; Secondary 65C40, 60J27, 93E15; 93E35.

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

The first author was supported by the Academy of Finland, projects no. 113826 and 118765, and by the Finnish Centre of Excellence in Analysis and Dynamics Research.

The second 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.

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 [9, 2, 5, 1, 15]; see also the review article [3]. 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, i.e. that I_n does not converge to $\pi(f)$ as $n \to \infty$ (see, e.g., [15] for an example). For practical considerations of the AM algorithm, the reader may consult [10, 16].

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 [2] 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 generalising 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 [15] 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 $\|x\| \to \infty$. In addition, we prove a central limit theorem, stating that $n^{-1/2} \sum_{k=1}^{n} [f(X_k) - \pi(f)]$ converges to a Gaussian random variable in distribution. Our main result (Theorem 13) 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 [12] to ensure the geometric convergence of a (non-adaptive) Metropolis process.

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 and a central limit theorem 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 21 by Meyn and Tweedie [14], 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 Notations

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,

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

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

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, 4] 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 $\mu(f):=\int f(x)\mu(\mathrm{d}x)$, and define $Pf(x):=\int f(y)P(x,\mathrm{d}y)$ 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 notations $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$.

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 $\epsilon \ge 0$.

(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 minorisation 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 minorisation 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

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

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

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

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

$$|H(s, x)| \le cn^\epsilon V^\beta(x).$$

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

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

The proof of Theorem 1 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 minorisation constants, then the speed of geometric convergence can get only polynomially worse.

Lemma 2. *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

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

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

Observe that the statement in Lemma 2 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 $s \in \cup_{n\geq 0} K_n$. 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.

Definition 3. *Suppose* $V : \mathbb{X} \to [1, \infty)$. *Given an increasing sequence of subsets* $K_n \subset \mathbb{S}$, $n \geq 1$, *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* $c \geq 1, \epsilon \geq 0$, *if for all* $s, s' \in K_n$, *we have*

$$\|f_s\|_V \leq c n^\epsilon \qquad \text{and} \qquad \|f_s - f_{s'}\|_V \leq c n^\epsilon |s - s'|.$$

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* $(c_3, \kappa_3\epsilon)$.

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

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

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* $(c_3, \kappa_3\epsilon)$. *In other words,*

$$\|\hat{f}_s\|_{V^r} + \|P_s\hat{f}_s\|_{V^r} \leq c_3 n^{\kappa_3\epsilon} \tag{8}$$

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

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

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

The part (i) follows easily from Lemma 2, since

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

Consider then (ii). The estimate (8) follows by the definition of $\hat{f}_s$ and Lemma 2,

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

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 2 and Assumption (A3), we have for all $s, s' \in K_n$ and $j \geq 0$

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

which gives that

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

Write then

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

By Lemma 2 and estimate (10) we have

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

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

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

$$\mathbb{E}\left[V^r(X_k)\right] \leq c_4^r k^{2r\epsilon} V^r(x_0) \tag{11}$$

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

where the constant c_4 depends only on c.

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

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

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

$$\mathbb{E}\left[V^r(X_k)\right] \leq \left(\mathbb{E}\left[V(X_k)\right]\right)^r \leq c_4^r k^{2r\epsilon} V^r(x_0).$$

Similarly, we have

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

by Jensen's inequality and the 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],

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

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

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

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 $c_* =$*

$c_*(c, p, \alpha, \beta, \xi) \geq 1$, *such that for all* $n \geq 1$,

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

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

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

whenever $\epsilon > 0$ *is small enough to ensure that* $\epsilon_* := \kappa_* \epsilon < \left[\frac{1}{2} \wedge \left(1 - \frac{1}{p}\right) \wedge (\xi - \alpha)\right]$, *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 $\|\hat{f}_s\|_{V^\alpha} + \|P_s \hat{f}_s\|_{V^\alpha} \leq c_3 \ell^{\kappa_3 \epsilon}$ for all $s \in K_\ell$. Since $\alpha p \in [0, 1]$, we can bound the martingale differences $\mathrm{d}M_\ell := M_\ell - M_{\ell-1}$ for $\ell \geq 1$ as follows,

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

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

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

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

$$\mathbb{E}|M_k|^p \leq c_p \mathbb{E}\left[\sum_{\ell=1}^k (|\mathrm{d}M_\ell|^p)^{2/p}\right]^{p/2} \leq c_p \sum_{\ell=1}^k \mathbb{E}|\mathrm{d}M_\ell|^p \leq \tilde{c} k^{(\kappa_3+2\alpha)p\epsilon+1} V^{\alpha p}(x_0).$$

The two cases combined give that

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

Now, by Corollary 23 of Birnbaum and Marshall's inequality in Appendix B,

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

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

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

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

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

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

By Proposition 4, $\|\hat{f}_s - \hat{f}_{s'}\|_{V^\alpha} \leq c_3 \ell^{\kappa_3\epsilon}|s - s'|$ 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 $|H(S_{\ell-1}, X_\ell)| \leq c\ell^\epsilon V^\beta(X_\ell)$, so

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

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

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

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

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

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

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

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

$$\begin{aligned}\mathbb{E}\left[\sup_{k\geq n} k^{-\xi p}|R_k^{(2)}|^p\right] &\leq 2^p\mathbb{E}\left[\sup_{k\geq n} k^{-\xi p}\left(|P_{S_0}\hat{f}_{S_0}(X_0)|^p + |P_{S_k}\hat{f}_{S_k}(X_k)|^p\right)\right] \\ &\leq 2^p\mathbb{E}\left[|P_{S_0}\hat{f}_{S_0}(X_0)|^p + \sup_{k\geq n} k^{-\xi p}|P_{S_k}\hat{f}_{S_k}(X_k)|^p\right] \\ &\leq \tilde{c} n^{(\kappa_3+2\alpha)p\epsilon + (\alpha-\xi)p}V^{\alpha p}(x_0).\end{aligned} \tag{21}$$

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

The proof of Theorem 1 follows as a straightforward application of Proposition 6.

Proof of Theorem 1. Let $\delta > 0$, and denote

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

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

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

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

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

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 $B := \cap_{m=1}^\infty \cup_{k=m}^\infty B_{n_k}^{(1/k)}$, the Borel-Cantelli lemma implies that $P(B^{\complement}) = 1$, and for all $\omega \in B^{\complement}$, (5) holds. □

Finally, we prove a central limit theorem in the lines of [1, Theorem 9], assuming one more condition holds, with the same constants $c \ge 1$ and $\epsilon \ge 0$ as (A1)–(A4).

(A5) There is a $\beta \in [0, 1/2]$ such that (A4) holds, and for all $n \ge 1$, $x \in \mathbb{X}$ and $s, s' \in K_n$,

$$|H(s,x) - H(s',x)| \le cn^\epsilon |s - s'| V^\beta(x).$$

Theorem 7. *Assume (A1)–(A5) hold. Let f be a function with $\|f\|_{V^\alpha} < \infty$ for some $\alpha \in (0, (1-\beta)/2)$. Assume $\epsilon < \kappa_{**}^{-1}[1/2 \wedge (1 - 2\alpha - \beta)]$ and $\sum_{k=1}^\infty k^{\kappa_{**}\epsilon - 1/2}\eta_k < \infty$, where $\kappa_{**} \ge 1$ is an independent constant. Furthermore, assume that S_k converges a.s. to some random variable S_∞, such that S_∞ belongs to the interior of K_N for some $N = N(\omega) < \infty$. Then,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n [f(X_k) - \pi(f)] \xrightarrow{n\to\infty} Z \tag{23}$$

in distribution, where Z is a random variable with characteristic function $\phi_Z(t) = \mathbb{E}e^{-\frac{1}{2}\sigma^2 t^2}$ with $\sigma^2 := \pi\big(\hat{f}_{S_\infty}^2 - (P_{S_\infty}\hat{f}_{S_\infty})^2\big) < \infty$.

Proof. Let $\kappa_{**} := 3\kappa_*^2$, where κ_* is the independent constant of Theorem 1. Consider again the martingale decomposition (14). As in the proof of Theorem 1, we can choose $p \in (1, (\alpha+\beta)^{-1})$ so that $\kappa_*\epsilon < (1-1/p)$, and let $\xi = 1/2$. Proposition 6 then implies that $n^{-1/2}(R_n^{(1)} + R_n^{(2)}) \to 0$ almost surely. So it suffices to show that $n^{-1/2}M_n \to Z$ in distribution. By the central limit theorem for martingales [11, Corollary 3.1], it is sufficient to show that for all $\varepsilon > 0$,

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[\mathrm{d}M_k^2 \;\middle|\; \mathcal{F}_{k-1}\right] \to \sigma^2 \qquad \text{and} \tag{24}$$

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[\mathrm{d}M_k^2 \mathbb{1}_{\{|\mathrm{d}M_k|\ge\varepsilon\sqrt{n}\}} \;\middle|\; \mathcal{F}_{k-1}\right] \to 0 \tag{25}$$

in probability, where $\mathrm{d}M_k := M_k - M_{k-1}$. Denote $g_s(x) := P_s\hat{f}_s^2(x) - (P_s\hat{f}_s)^2$, and notice that

$$\mathbb{E}\left[\mathrm{d}M_k^2 \;\middle|\; \mathcal{F}_{k-1}\right] = g_{S_{k-1}}(X_{k-1}).$$

In the present setting, Proposition 4 yields that the families $\{\hat{f}_s\}_{s\in\mathbb{S}}$ and $\{P_s\hat{f}_s\}_{s\in\mathbb{S}}$ are (K_n, V^α)-polynomially Lipschitz with constants $(c_3, \kappa_3\epsilon)$, implying that $\{\hat{f}_s^2\}_{s\in\mathbb{S}}$ and $\{(P_s\hat{f}_s)^2\}_{s\in\mathbb{S}}$ are $(K_n, V^{2\alpha})$-polynomially Lipschitz with constants $(2c_3^2, 2\kappa_3\epsilon)$. Since $\kappa_* > \kappa_3 \vee \kappa_2$, we obtain that $\{g_s\}_{s\in\mathbb{S}}$ is $(K_n, V^{2\alpha})$-polynomially Lipschitz with constants $(\tilde{c}, 3\kappa_*\epsilon)$ for some $\tilde{c} \ge 1$. We can choose again $p \in (1, (2\alpha+\beta)^{-1})$ such that $3\kappa_*^2\epsilon < (1 - 1/p)$, and apply Proposition 6 to obtain

$$\frac{1}{n}\sum_{k=1}^{n}\left(\mathbb{E}\left[\mathrm{d}M_k^2 \;\middle|\; \mathcal{F}_{k-1}\right] - \pi(g_{S_{k-1}})\right) \to 0$$

almost surely. Since $S_k \to S_\infty$ almost surely, and S_∞ is in the interior of K_N, there is an a.s. finite N' such that $S_k \in K_{N'}$ for all $k \ge 1$, and

$$|\pi(g_{S_k}) - \pi(g_{S_\infty})| \le \|g_{S_k} - g_{S_\infty}\|_{V^\alpha}\pi(V^\alpha) \le \tilde{c}|S_k - S_\infty| \to 0.$$

That is, $\pi(g_{S_k}) \to \pi(g_{S_\infty})$, and hence

$$\frac{1}{n}\sum_{k=0}^{n-1}\pi(g_{S_k}) \to \pi(g_{S_\infty}) = \sigma^2 < \infty$$

almost surely. This yields (24).

Consider then (25). Applying Lemma 24 in Appendix B, we obtain that

$$\mathbb{E}\left[\mathrm{d}M_k^2 \mathbb{1}_{\{|\mathrm{d}M_k|\ge\varepsilon\sqrt{n}\}} \;\middle|\; \mathcal{F}_{k-1}\right] \le 4\mathbb{E}\big[\hat{f}_{S_{k-1}}^2(X_k)\mathbb{1}_{\{|\hat{f}_{S_{k-1}}(X_k)|\ge\varepsilon\sqrt{n}/2\}} \;\big|\; \mathcal{F}_{k-1}\big].$$

It follows for all $\varepsilon, L > 0$ and for sufficiently large $n \ge 1$ that

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[\mathrm{d}M_k^2 \mathbb{1}_{\{|\mathrm{d}M_k|\ge\varepsilon\sqrt{n}\}} \;\middle|\; \mathcal{F}_{k-1}\right] \le \frac{4}{n}\sum_{k=1}^{n} h_{S_{k-1}}^{(L)}(X_{k-1}) \tag{26}$$

where

$$h_s^{(L)}(x) := \int P_s(x,\mathrm{d}y)\hat{f}_s^2(y) \sup_{s'\in K_{N'}} \mathbb{1}_{\{|\hat{f}_{s'}(y)|>L\}}.$$

and where the supremum can be taken with respect to a countable dense subset of $K_{N'}$ to ensure measurability. As before, one checks that for all $L > 0$ the family $\{h_s^{(L)}\}_{s\in\mathbb{S}}$ is $(K_n, V^{2\alpha})$-polynomially Lipschitz with constants $(\tilde{c}, 3\kappa_*\epsilon)$, and hence the right hand side of (26) converges almost surely to

$$4\pi\big(h_{S_\infty}^{(L,m)}\big) = 4\int \hat{f}_{S_\infty}^2(x) \sup_{s\in K_{N'}} \mathbb{1}_{\{|\hat{f}_s(x)|>L\}}\pi(x)\mathrm{d}x \xrightarrow{L\to\infty} 0$$

by monotone convergence. □

Remark 8. *Theorem 7 assumes that the adaptation parameter S_n converges to some finite limit S_∞. The convergence of S_n in our sequentially constrained adaptive MCMC, in general, is out of the scope of this paper, and might require additional conditions on the adaptation mechanism. However, in Section 5, we see that in the case of Adaptive Metropolis this can be verified fairly easily.*

Remark 9. *The constraint functions σ_n of our framework can be defined more generally by allowing σ_n to have additional dependence on ω in a $\mathcal{F}_{n-1}$-measurable (predictable) manner. The proofs above do not need to be modified to cover this generalisation. Indeed, we employ a different definition for σ_n in the proof of the central limit theorem for the adaptive Metropolis process in Section 5.*

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 $(\|X_n\|)_{n\geq 1}$, 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 10. *Suppose that there is $V : \mathbb{X} \to [1,\infty)$ such that the bound*

$$P_s V(x) \leq V(x) + b \tag{27}$$

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

$$\|x\| \geq u \implies V(x) \geq r(u) \tag{28}$$

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

$$\sup_{u\geq 1} \frac{u^p}{r(u)} \leq c. \tag{29}$$

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

$$\|X_n\| \leq A n^\epsilon.$$

Proof. To start with, (27) implies for $n \geq 1$

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

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

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

where we use Markov's inequality, and $c = c(3/\epsilon) < \infty$ is from the application of (29). □

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

Lemma 11. *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 (28) *applies with a function $r(u) := ce^{\gamma u}$ for some $\gamma, c > 0$, satisfying* (29)*.*

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

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

We have that

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

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

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 $n \geq 0$,

$$X_{n+1} \sim P_{\theta\Sigma_n}(X_n, \cdot\,) \tag{30}$$

$$\Sigma_{n+1} = \begin{cases} v_0, & 0 \leq n \leq N_b - 1 \\ \mathrm{Cov}(X_0, \ldots, X_n) + \kappa I, & n \geq N_b \end{cases} \tag{31}$$

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

$$P_v(x, A) := \mathbb{1}_A(x) \left[1 - \int \left(1 \wedge \frac{\pi(y)}{\pi(x)}\right) q_v(y-x)\mathrm{d}y\right] + \int_A \left(1 \wedge \frac{\pi(y)}{\pi(x)}\right) q_v(y-x)\mathrm{d}y \tag{32}$$

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 12), we consider a slight modification of the AM chain. Firstly, we do not consider a burn-in period, i.e. let $N_b = 0$, and let $\Sigma_0 \geq \kappa I$. Instead of (31), we construct Σ_n recursively for $n \geq 1$ as

$$\Sigma_n = \frac{n}{n+1}\Sigma_{n-1} + \frac{1}{n+1}\left[(X_n - \overline{X}_{n-1})(X_n - \overline{X}_{n-1})^T + \kappa I\right] \tag{33}$$

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

Remark 12. *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*

$$\Sigma_n = \frac{n-1}{n}\Sigma_{n-1} + \frac{1}{n+1}\left[(X_n - \overline{X}_{n-1})(X_n - \overline{X}_{n-1})^T + \kappa I\right] \tag{34}$$

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 (33) *for simpler notations. Also, from a practical point of view, observe that* (33) *differs from* (34) *by a factor smaller than $n^{-2}\Sigma_{n-1}$ whence it is mostly a matter of taste whether to use* (33) *or* (34).

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 $\|A\|^2 := \mathrm{trace}(A^T A)$, respectively. The Metropolis kernel P_s is defined as in (32), 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 13. *Assume π is positive, bounded, bounded from below on compact sets, differentiable, and*

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|^\rho} \cdot \nabla \log \pi(x) = -\infty \tag{35}$$

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

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|} \cdot \frac{\nabla\pi(x)}{\|\nabla\pi(x)\|} < 0. \tag{36}$$

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 $0 \leq \alpha < 1$,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{n\to\infty} \pi(f) \tag{37}$$

almost surely. If, in addition, $\alpha < 1/2$,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} [f(X_k) - \pi(f)] \xrightarrow{n\to\infty} N(0, \sigma^2) \tag{38}$$

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

Remark 14. *If the conditions of Theorem 13 are satisfied, the function* $V(x)$ *grows faster than an exponential, and hence* (37) *and* (38) *hold for exponential moments. In particular, they hold for power moments, i.e. 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 13 is postponed to the end of this section. We start by a simple lemma bounding the growth rate of the AM chain.

Lemma 15. *If the conditions of Proposition 10 are satisfied for an AM chain, then for any* $\epsilon > 0$, *there is an a.s. finite* $A = A(\omega, \epsilon)$ *such that*

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

Proof. Since the AM recursion is a convex combination, this is a straightforward corollary of Proposition 10. □

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 18 is similar to the results obtained in [12, 17], with the exception that we have a common minorisation 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 16. *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 $\epsilon > 0$, *define* $B_\epsilon := \{ru : u \in S^d,\ g(u) < r \leq g(u) + \epsilon\}$. *Then, for all* $\tilde{\epsilon} > 0$, *there is a* $\tilde{b} = \tilde{b}(\tilde{\epsilon}) \in (0, \infty)$ *such that for all* $0 < \epsilon < \tilde{\epsilon}$ *and for all* $\lambda \geq 3\epsilon$, *it holds that*

$$|E \cap B_\epsilon| \leq \left|\left(E \oplus \overline{B}(0, \lambda)\right) \cap A\right|$$

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

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

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

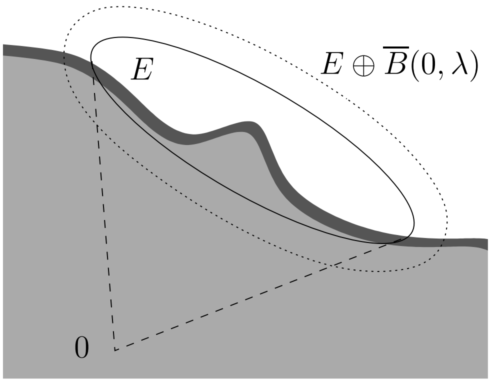

FIGURE 1. Illustration of the boundary estimate. The set A is in light grey, and the set B_ϵ in dark grey.

$(E \cap B_\epsilon) \oplus \overline{B}(0,\lambda)$, and we can estimate

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

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

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

whenever $\epsilon \leq c_d y$. □

Lemma 17. *Let* $f(x) := xe^{-\frac{x^2}{2}}$. *For any* $0 < \epsilon < 1/8$, *the following estimates hold*

$$2f(x+\epsilon) - f(x) \geq \frac{x}{8}, \qquad \textit{for all} \qquad 0 < x \leq \frac{1}{2}, \quad \textit{and}$$

$$\int_0^\infty \big([2f(x+\epsilon) - f(x)] \wedge 0\big)\mathrm{d}x \geq -e^{-c\epsilon^{-2}}$$

for some constant $c > 0$.

Proof. We can write

$$2f(x+\epsilon) - f(x) = e^{-\frac{x^2}{2}}\left[2(x+\epsilon)e^{-x\epsilon - \frac{\epsilon^2}{2}} - x\right]$$

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

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

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

$$2f(x+\epsilon) - f(x) \geq \frac{1}{4} x e^{-\frac{x^2}{2}} \geq \frac{x}{8}$$

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

$$\int_0^\infty \big([2f(x+\epsilon) - f(x)] \wedge 0\big) \mathrm{d}x \geq -\int_{x^*}^\infty x e^{-\frac{x^2}{2}} \mathrm{d}x = -e^{-c\epsilon^{-2}}$$

with $c = 1/32$. □

Proposition 18. *Assume that π satisfies the conditions in Theorem 13 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 (32) and for all $v \in \mathcal{C}^d$ with all eigenvalues greater than $\kappa > 0$, it holds that*

$$P_v V(x) \leq \lambda_v V(x) + b \mathbb{1}_C(x), \qquad \forall x \in \mathbb{X} \tag{39}$$

$$P_v(x, B) \geq \delta_v \nu(B) \qquad \forall x \in C, \forall B \subset \mathbb{X} \tag{40}$$

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 25 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 parameterise $A_x = \{ru : u \in S^d,\ 0 \leq r \leq g(u)\}$ 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 (39). We may compute

$$\begin{aligned} \tau_v := 1 - \frac{P_v V(x)}{V(x)} = &\int_{A_x} \left(1 - \sqrt{\frac{\pi(x)}{\pi(y)}}\right) q_v(y-x) \mathrm{d}y \\ &- \int_{R_x} \sqrt{\frac{\pi(y)}{\pi(x)}} \left(1 - \sqrt{\frac{\pi(y)}{\pi(x)}}\right) q_v(y-x) \mathrm{d}y. \end{aligned} \tag{41}$$

In what follows, unless explicitly stated, we assume $\|x\| \geq M(R+1)$. Denote $\epsilon_x := \|x\|^{-\alpha} < 1$, where $\alpha = (\rho-1)/2 > 0$. Define $\tilde{A}_x := \{ru : u \in S^d,\ 0 \leq r \leq g(u) - \epsilon_x\} \subset A_x$ and $\tilde{R}_x := \{ru : u \in S^d,\ r \geq g(u) + \epsilon_x\} \subset R_x$. From (41), we can estimate

$$\tau_v \geq \int \left[\left(1 - \sqrt{\frac{\pi(x)}{\pi(y)}}\right) \mathbb{1}_{\tilde{A}_x}(y) - \frac{1}{4}\mathbb{1}_{R_x \setminus \tilde{R}_x}(y)\right] q_v(y-x)\mathrm{d}y \\ - \sup_{z\in\mathbb{R}^d} q_v(z-x) \int_{\tilde{R}_x} \sqrt{\frac{\pi(y)}{\pi(x)}}\mathrm{d}y. \tag{42}$$

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 $\epsilon_x \leq a \leq \|y\|$. Assume for a moment $\|z\| \geq R$. Then, h is decreasing on the line segment from z to y, and we can estimate

$$\frac{\pi(x)}{\pi(z)} = \frac{\pi(y)}{\pi(z)} = e^{h(y)-h(z)} = e^{\int_{\|y\|-a}^{\|y\|} \frac{y}{\|y\|}\cdot\nabla h(t\frac{y}{\|y\|})\mathrm{d}t} \leq e^{\int_{\|y\|-\epsilon_x}^{\|y\|} \frac{y}{\|y\|}\cdot\nabla h(t\frac{y}{\|y\|})\mathrm{d}t} \\ \leq e^{-\epsilon_x(\|y\|-\epsilon_x)^{\rho-1}} \leq e^{-\epsilon_x\|x\|^{\rho-1}/(2M)^{\rho-1}} = e^{-\|x\|^{\alpha}/(2M)^{\rho-1}}$$

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,

$$\frac{\pi(x)}{\pi(z)} = \frac{\pi(y)}{\pi(z')}\frac{\pi(z')}{\pi(z)} \leq e^{-\|x\|^{\alpha}/(2M)^{\rho-1}} \frac{\sup_{\|w\|\leq R}\pi(w)}{\inf_{\|w\|\leq R}\pi(w)} \leq \frac{1}{4} \tag{43}$$

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

$$\int f(z+x)q_v(z)\mathrm{d}x = \frac{c_d}{\sqrt{\det(v)}} \int_0^1 \int_{\{e^{-\frac{1}{2}z^T v^{-1} z} \geq t\}} f(z+x)\mathrm{d}z\mathrm{d}t \\ = \frac{c_d}{\sqrt{\det(v)}} \int_0^\infty \int_{E_u} f(y)\mathrm{d}y u e^{-\frac{u^2}{2}}\mathrm{d}u$$

where $c_d = (2\pi)^{-d/2}$ and $E_u := \{z + x : z^T v^{-1} z \leq u^2\}$. Consequently, for $\|x\| \geq R_2$, we can estimate the first term of (42) from below by

$$\int_0^\infty \left(\frac{|E_u \cap \tilde{A}_x|}{2} - \frac{|E_u \cap (R_x \setminus \tilde{R}_x)|}{4}\right) u e^{-\frac{u^2}{2}}\mathrm{d}u \\ \geq \frac{1}{4}\int_0^\infty 2|E_{u+a}\cap\tilde{A}_x|(u+a)e^{-\frac{(u+a)^2}{2}} - |E_u \cap (R_x \setminus \tilde{R}_x)| u e^{-\frac{u^2}{2}}\mathrm{d}u \\ \geq \frac{1}{4}\int_0^\infty 2\big|\big(E_u \oplus \overline{B}(0,\kappa^{1/2}a)\big)\cap\tilde{A}_x\big|(u+a)e^{-\frac{(u+a)^2}{2}} - |E_u \cap B_\epsilon| u e^{-\frac{u^2}{2}}\mathrm{d}u$$

for any $a \geq 0$, 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 $\tilde{A}_x = \{ru : u \in S^d,\, 0 \leq r \leq \tilde{g}(u)\}$ where $\tilde{g}(u) = g(u) - \epsilon_x$, we obtain that $R_x \setminus \tilde{R}_x \subset \{ru : u \in S^d,\, \tilde{g}(u) \leq r \leq \tilde{g}(u) + 2\epsilon_x\} =: B_\epsilon$. We set $a = 6\kappa^{-1/2}\epsilon_x$ and apply Lemma 16 with the choice

$\epsilon = 2\epsilon_x$ and $\lambda = 6\epsilon_x$,

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

by Lemma 17, 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 $\epsilon_x \to 0$ 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 $|\overline{B}(x, \kappa^{1/2}/4) \cap A_x| \geq c_3 > 0$ (see the proof of Theorem 4.3 in [12]). So, for large enough $\|x\|$, there is a $c_4 > 0$ so that $|E_{1/4} \cap \tilde{A}_x| \geq c_4$. To sum up, by choosing R_3 to be sufficiently large, we obtain that the first part of (42) is at least $c_5(\det(v))^{-1/2}$ for all $\|x\| \geq R_3$, with a $c_5 > 0$.

Next, we turn to the second term of (42). We obtain by polar integration that

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

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

$$
\begin{aligned}
\int_{w+\epsilon_x}^\infty e^{-\frac{1}{4} \int_w^r t^{\rho-1} \mathrm{d}t} \mathrm{d}r \sup_{r \geq w+\epsilon_x} T(w,r) &\leq \int_w^\infty e^{-\frac{w^{\rho-1}}{4}(r-w)} \mathrm{d}r \sup_{r \geq w+\epsilon_x} T(w,r) \\
&\leq 4M^{\rho-1} \|x\|^{1-\rho} \sup_{r \geq w+\epsilon_x} T(w,r).
\end{aligned}
$$

for any $w \geq M^{-1}\|x\|$. Suppose first $w + \epsilon_x \leq r \leq 2w$, then

$$
T(w,r) \leq (2w)^{d-1} e^{-\frac{1}{4}\epsilon_x w^{\rho-1}} \leq (2M)^{d-1} \|x\|^{d-1} e^{-\frac{1}{4} M^{1-\rho} \|x\|^\alpha} \leq c_6
$$

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

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

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

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 $P_v V(x) \leq \lambda_v V(x)$ holds for $x \notin C$ with $\lambda_v := 1 - c_8 \det(v)^{-1/2}$ satisfying $(1-\lambda_v)^{-1} \leq c_8^{-1} \det(v)^{1/2}$. For $x \in C$, we have by (41) that $P_v V(x) \leq 2V(x) \leq 2\sup_{z \in C} V(z) \leq b < \infty$, so (39) 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 minorisation condition (40) for $x \in C$,

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

So (40) holds with $\delta_v := c_9 \det(v)^{-1/2}$ for some $c_9 > 0$. Finally, the claim holds with $c := c_8^{-1} \vee c_9^{-1}$. □

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

Proof of Theorem 13. We start by verifying the strong law of large numbers (37). Fix $t \geq 1$ and consider first the constrained process $(X_n^{(t)}, S_n^{(t)})_{n\geq 0}$ which is defined as the AM chain, but with the constraint sets $K_n^{(t)}$ defined as $K_n^{(t)} := \{s \in \mathbb{S} : |s| \leq tn^{\epsilon'}\}$, with $\epsilon' = \epsilon/(2d)$, and $\epsilon \in (0, \kappa_*^{-1}[(1/2) \wedge (1-\alpha)])$, where κ_* is the independent constant of Theorem 1.

We check that the assumptions (A1)–(A4) are satisfied by the constrained process $(X_n^{(t)}, S_n^{(t)})_{n\geq 0}$ for all $t \geq 1$. The condition (A1) is satisfied by construction of the Metropolis kernels P_s. Since $\det(v) \leq \|v\|^d$, Proposition 18 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 as in [4], we have that

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

for any $\beta \in (0, 1/2]$ by Lemma 11, 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 $\epsilon < \kappa_*^{-1}[(1/2)\wedge(1-\alpha-\beta)]$. Clearly, $\sum_k k^{\kappa_*\epsilon-1}\eta_k < \sum_k k^{\kappa_*\epsilon-2} < \infty$, so all the conditions of Theorem 1 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 15 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 1, we can use the Borel-Cantelli lemma to deduce that (37) holds almost surely.

We finally verify the central limit theorem (38). Define $m_\pi := \int x\pi(x)\mathrm{d}x$ and $\tilde{v}_\pi := \int xx^T\pi(x)\mathrm{d}x - m_\pi m_\pi^T + \kappa I$ as the mean and (modified) covariance of the distribution π, which are finite, as observed in Remark 14. In addition, as noted in Remark 12, $S_n \to s_\pi := (m_\pi, \tilde{v}_\pi)$ almost surely by (37). Therefore, if one denotes $A_t := \{\sup_{n\geq 0} |S_n| \leq t\}$, then $\mathbb{P}(A_t) \to 1$ as $t \to \infty$.

Fix $t > |s_\pi| \vee |s_0|$. Define the sets $\tilde{K}_n^{(t)} := \tilde{K}^{(t)} := K_1^{(t)}$ for all $n \geq 0$ and let

$$\tilde{\sigma}_n^{(t)}(s, s') := \begin{cases} s + s', & \text{if } S_{k-1} + \eta_k H(S_{k-1}, X_k) \in \operatorname{int} \tilde{K}^{(t)} \text{ for all } 1 \leq k \leq n \\ s, & \text{otherwise} \end{cases}$$

where $\operatorname{int} \tilde{K}^{(t)}$ stands for the interior of $\tilde{K}^{(t)}$. Define the constrained process $(\tilde{X}_n^{(t)}, \tilde{S}_n^{(t)})_{n\geq 0}$ following the framework of Section 2 with constraint functions $\tilde{\sigma}_n^{(t)}$. Here one observes that our constraints $\tilde{\sigma}_n^{(t)}$ correspond to stopping the adaptation at the time of possible first exit from the interior of $\tilde{K}^{(t)}$, whence Remark 9 applies in the present situation. With this definition, the assumptions (A1)–(A4) are satisfied with some $\tilde{c} = \tilde{c}(t) \geq 1$ and $\epsilon = 0$, and similarly as above, we obtain for $s, s' \in K_n^{(t)}$

$$\begin{aligned}
&|H(s, x) - H(s', x)| \\
&\quad \leq \|s^{(m)} - s'^{(m)}\| + \|s^{(v)} - s'^{(v)}\| \\
&\qquad + \|(x - s^{(m)})(x - s^{(m)})^T - (x - s'^{(m)})(x - s'^{(m)})^T\| \\
&\quad \leq \big[1 + 2\,\|x\| + 2(\|s^{(m)}\| \vee \|s'^{(m)}\|)\big]\,\|s^{(m)} - s'^{(m)}\| + \|s^{(v)} - s'^{(v)}\| \\
&\quad \leq \tilde{c}t(1 \vee \|x\|)|s - s'|
\end{aligned}$$

and then that

$$\|H(s, x) - H(s', x)\|_{V^\beta} \leq \tilde{c}t|s - s'| \sup_{x\in\mathbb{R}^d} \frac{1 \vee \|x\|}{V^\beta(x)} \leq \tilde{c}t|s - s'|$$

for $s, s' \in K_n^{(t)}$ establishing (A5) for any $\beta > 0$.

The process $(\tilde{X}_n^{(t)}, \tilde{S}_n^{(t)})_{n\geq 0}$ coincides with the AM chain $(X_n, S_n)_{n\geq 0}$ in A_t, in which the adaptation parameters $\tilde{S}_n^{(t)}$ converge almost surely to $s_\pi \in \operatorname{int} \tilde{K}^{(t)}$. In the complement of A_t, the parameters S_n converge almost surely to some $S_\infty \in \operatorname{int} \tilde{K}^{(t)}$. We can apply Theorem 7 to deduce that

$$\tilde{Y}_n^{(t)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \Big[f(\tilde{X}_k^{(t)}) - \pi(f)\Big] \xrightarrow{n\to\infty} \tilde{Z}^{(t)}$$

in distribution, where $\tilde{Z}^{(t)}$ is a random variable with the characteristic function $\phi_{\tilde{Z}^{(t)}}(u) = \mathbb{E}e^{-\frac{1}{2}\tilde{\sigma}_t^2 u^2}$, where $\tilde{\sigma}_t^2$ is finite almost surely, and equals to σ^2 in A_t. Let $Z \sim N(0, \sigma^2)$, i.e. $\phi_Z(u) = e^{-\frac{1}{2}\sigma^2 u^2}$. For fixed $u \in \mathbb{R}$, we have

$$\begin{aligned}
|\phi_Z(u) - \phi_{\tilde{Z}^{(t)}}(u)| &\leq \phi_Z(u) \int \Big|1 - e^{\frac{1}{2}(\sigma^2 - \tilde{\sigma}_t^2)u^2}\Big| \mathrm{d}\mathbb{P} \\
&\leq \phi_Z(u)[1 - \mathbb{P}(A_t)]\big[1 \vee (e^{\frac{\sigma^2}{2}} - 1)\big]
\end{aligned}$$

so the characteristic functions $\phi_{\tilde{Z}^{(t)}}$ converge pointwise to ϕ_Z, and hence $\tilde{Z}^{(t)} \xrightarrow{t\to\infty} Z$ in distribution.

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be bounded and continuous, and denote the probability measures induced by random variables as $\mu_X(A) := \mathbb{P}(X \in A)$. We can choose a non-decreasing sequence $(t_n)_{n\geq 1}$ of positive numbers such that $t_n \to \infty$ and

$$|\mu_{\tilde{Y}_n^{(t_n)}}(\varphi) - \mu_{Z^{(t_n)}}(\varphi)| \xrightarrow{n\to\infty} 0.$$

Since $\tilde{Y}_n^{(t_n)}$ is equal to $Y_n := n^{-1/2}\sum_{k=1}^n [f(X_k) - \pi(f)]$ in A_{t_n}, we have that

$$|\mu_{Y_n}(\varphi) - \mu_{\tilde{Y}_n^{(t_n)}}(\varphi)| \leq [1 - \mathbb{P}(A_{t_n})] \sup_{x\in\mathbb{R}} |\varphi(x)| \xrightarrow{n\to\infty} 0.$$

We conclude that $|\mu_{Y_n}(\varphi) - \mu_Z(\varphi)| \to 0$ as $n \to \infty$, and (38) holds. □

Remark 19. *Since $\epsilon > 0$ can be selected arbitrarily small in the proof of Theorem 13, it is only required for* (37) *to hold that the adaptation weights $\eta_n \in (0,1)$ are decreasing and that $\sum_k k^{\tilde{\epsilon}-1}\eta_k < \infty$ holds for some $\tilde{\epsilon} > 0$. In particular, one can choose $\eta_n := (n+1)^{-\gamma}$ for any $\gamma > 0$.*

Remark 20. *The condition* (35) *implies the super-exponential decay of the tails of π*

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|} \cdot \nabla \log \pi(x) = -\infty. \tag{44}$$

This condition, with the contour regularity condition (36)*, are common conditions to ensure geometric ergodicity of a random-walk Metropolis algorithm, and many standard distributions fulfil them [12]. The decay condition* (35) *is only slightly more stringent than* (44)*.*

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. F. Atchadé and J. S. Rosenthal. On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, 11(5):815–828, 2005.

[6] P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15(1A):700–738, 2005.

[7] Z. W. Birnbaum and A. W. Marshall. Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *The Annals of Mathematical Statistics*, 32(2):687–703, Sept. 1961.

[8] A. Dvoretzky. Asymptotic normality for sums of dependent random variables. In *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, volume 2, pages 513–535. Univ. of California Press, 1972.

[9] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.
[10] H. Haario, M. Laine, M. Lehtinen, E. Saksman, and J. Tamminen. Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(3):591–607, 2004.
[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] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, June 1953.
[14] S. P. Meyn and R. L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, 4(4):981–1011, 1994.
[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. *J. Comput. Graph. Statist.*, 18(2):349–367, 2009.
[17] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.

Appendix A. Proof of Lemma 2

We provide a restatement of a part of a theorem by Meyn and Tweedie [14] before proving Lemma 2. For a more recent work on quantitative convergence bounds, we refer to [6].

Theorem 21. *Suppose that the following drift and minorisation 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 $\rho > \vartheta = 1 - \widetilde{M}^{-1}$, *for*

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

defined in terms of

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

and the bound

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

Proof. [14, Theorem 2.3]. □

Proof of Lemma 2. Observe that $P_s V(x) = \mathbb{E}\left[V(X_{n+1}) \mid X_n = x, S_n = s\right]$, 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 $\tilde{\lambda}_n := \lambda_n^r \le (1 - c^{-1}n^{-\epsilon})^r \le 1 - rc^{-1}n^{-\epsilon}$ implying

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

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

Now, we can apply Theorem 21, where we can estimate the constants

$$\gamma_n = \delta_n^{-2}\left[4\tilde{b}_n + 2\delta_n \tilde{\lambda}_n \tilde{b}_n\right] \le (cn^{\epsilon})^2 6(\tilde{c}n^{\epsilon}) = a_1 n^{3\epsilon}$$
$$\check{b}_n = \tilde{b}_n + \gamma_n \le (\tilde{c} + a_1)n^{3\epsilon} \le a_2 n^{3\epsilon}$$

and consequently

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

Moreover,

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

and then

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

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

$$1 - \vartheta_n = \widetilde{M}_n^{-1} \ge a_5^{-1} n^{-23\epsilon}$$

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

$$\rho_n - \vartheta_n = 1 - \rho_n = \frac{1}{2}(1 - \vartheta_n) \ge \frac{1}{2} c_9^{-1} n^{-23\epsilon} = (a_6 n^{23\epsilon})^{-1}.$$

Finally, from Theorem 21, one obtains the bound

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

where

$$(1 - \rho_n)^{-1} \le a_6 n^{23\epsilon}$$
$$L_n = (1 + \gamma_n)\frac{\rho_n}{\rho_n - \vartheta_n} \le (1 + a_1 n^{3\epsilon})(a_6 n^{23\epsilon}) \le a_7 n^{26\epsilon}$$

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

Appendix B. Some General Inequalities

Theorem 22 (Birnbaum and Marshall). *Let $(X_k)_{k=1}^n$ be random variables, such that*

$$\mathbb{E}\left[|X_k| \mid \mathcal{F}_{k-1}\right] \ge \psi_k |X_{k-1}|,$$

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

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

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

$$\mathbb{P}\left(\max_{1\le k\le n} a_k |X_k| \ge 1\right) \le \sum_{k=1}^n (b_k^p - \psi_{k+1}^p b_{k+1}^p)\mathbb{E}|X_k|^p.$$

Proof. [7, Theorem 2.1]. □

Corollary 23. *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 non-increasing sequence. If $p \ge 1$ is such that $\mathbb{E}|M_k|^p < \infty$ for all $1 \le k \le n$, then for $1 \le m \le n$,*

$$\mathbb{P}\left(\max_{m\le k\le n} a_k |M_k| \ge 1\right) \le a_n^p \mathbb{E}|M_n|^p + \sum_{k=m}^{n-1} (a_k^p - a_{k+1}^p)\mathbb{E}|M_k|^p.$$

Proof. By Jensen's inequality,

$$\mathbb{E}\left[|M_k| \mid \mathcal{F}_{k-1}\right] \ge \left|\mathbb{E}\left[M_k \mid \mathcal{F}_{k-1}\right]\right| = |M_{k-1}|.$$

Define $\psi_k := 1$ for $1 \le k \le n$, and $\tilde{a}_k := a_m$ for $1 \le k \le m$ and $\tilde{a}_k := a_k$ for $m < k \le n$. The result follows from Theorem 22. □

The following lemma is a conditional version of [8, Lemma 3.3], and was stated also in [1, Lemma 10].

Lemma 24 (Dvoretzky). *Let X be a square integrable random variable and $\mathcal{G}$ a σ-algebra on a probability space. Then, for every $\varepsilon > 0$,*

$$\mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{G}])^2 \mathbb{1}_{\{|X - \mathbb{E}[X \mid \mathcal{G}]| \ge 2\varepsilon\}} \;\middle|\; \mathcal{G}\right] \le 4\mathbb{E}\left[X^2 \mathbb{1}_{\{|X|\ge\varepsilon\}} \;\middle|\; \mathcal{G}\right].$$

Proof. Notice that $\mathbb{1}_{\{|X-\mathbb{E}[X \mid \mathcal{G}]|\ge 2\varepsilon\}} \le \mathbb{1}_{\{|\mathbb{E}[X \mid \mathcal{G}]|\ge\varepsilon\}} + \mathbb{1}_{\{|X|\ge\varepsilon, |\mathbb{E}[X \mid \mathcal{G}]|<\varepsilon\}}$. We can estimate

$$\begin{aligned}
&\mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{G}])^2 \mathbb{1}_{\{|\mathbb{E}[X \mid \mathcal{G}]|\ge\varepsilon\}} \;\middle|\; \mathcal{G}\right] \\
&\quad = \mathbb{E}\left[(X^2 - \mathbb{E}[X \mid \mathcal{G}]^2) \mathbb{1}_{\{|\mathbb{E}[X \mid \mathcal{G}]|\ge\varepsilon\}} \;\middle|\; \mathcal{G}\right] \\
&\quad \le \mathbb{E}\left[(X^2 - \varepsilon^2) \vee 0 \;\middle|\; \mathcal{G}\right] = \mathbb{E}\left[(X^2 - \varepsilon^2) \mathbb{1}_{\{|X|\ge\varepsilon\}} \;\middle|\; \mathcal{G}\right].
\end{aligned} \tag{45}$$

Similarly, we obtain

$$\begin{aligned}
&\mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{G}])^2 \mathbb{1}_{\{|X|\ge\varepsilon, |\mathbb{E}[X \mid \mathcal{G}]|<\varepsilon\}} \;\middle|\; \mathcal{G}\right] \\
&\quad \le \mathbb{E}\left[(X^2 + 2\varepsilon|X| + \varepsilon^2) \mathbb{1}_{\{|X|\ge\varepsilon, |\mathbb{E}[X \mid \mathcal{G}]|<\varepsilon\}} \;\middle|\; \mathcal{G}\right] \\
&\quad \le \mathbb{E}\left[(3X^2 + \varepsilon^2) \mathbb{1}_{\{|X|\ge\varepsilon\}} \;\middle|\; \mathcal{G}\right].
\end{aligned} \tag{46}$$

Summing (45) and (46) concludes the proof. □

Appendix C. Contour Surface Containment

Lemma 25. *Suppose $A \subset \mathbb{R}^d$ is a smooth surface parameterised 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

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

and hence we have $|\frac{\mathrm{d}}{\mathrm{d}\theta} \log r(\theta))| = |r'(\theta)/r(\theta)| \leq \tan \alpha_0$ 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. □

Eero Saksman, Department of Mathematics and Statistics, University of Helsinki, P.O.Box 68, FI-00014 University of Helsinki, Finland
E-mail address: eero.saksman@helsinki.fi

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/