Available online at www.sciencedirect.com
SciVerse ScienceDirect
Stochastic Processes and their Applications 121 (2011) 2839–2860
stochastic processes and their applications
www.elsevier.com/locate/spa

On the stability and ergodicity of adaptive scaling Metropolis algorithms

Matti Vihola*

Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35, FI-40014, Finland

Received 7 May 2010; received in revised form 10 August 2011; accepted 10 August 2011
Available online 23 August 2011

Abstract

The stability and ergodicity properties of two adaptive random walk Metropolis algorithms are considered. Both algorithms adjust the scaling of the proposal distribution continuously based on the observed acceptance probability. Unlike the previously proposed forms of the algorithms, the adapted scaling parameter is not constrained within a predefined compact interval. The first algorithm is based on scale adaptation only, while the second one also incorporates covariance adaptation. A strong law of large numbers is shown to hold assuming that the target density is smooth enough and has either compact support or super-exponentially decaying tails.

MSC: primary 65C40; secondary 60J27; 93E15; 93E35

Keywords: Adaptive Markov chain Monte Carlo; Law of large numbers; Metropolis algorithm; Stability; Stochastic approximation

1. Introduction

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

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

* Tel.: +358 40 805 3452; fax: +358 14 260 2701.
E-mail address: matti.vihola@iki.fi.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2011.08.006

where π is a probability density function (see, e.g., [9,15,18]). The method is based on a Markov chain $(X_n)_{n\geq 1}$ that can be simulated in practice, and for which the ergodic averages $I_n := n^{-1}\sum_{k=1}^{n} f(X_k)$ converge to the integral I as $n \to \infty$. Such a chain can be constructed, for example, as follows. Assume q is a standard Gaussian probability density in $\mathbb{R}^d$, and let $X_1 \equiv x_1$ for some fixed point $x_1 \in \mathbb{R}^d$. For $n \geq 2$, recursively,

(S1) set $Y_n := X_{n-1} + \theta\Sigma^{1/2} W_n$, where W_n are independent random vectors distributed according to q, and

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

For any scale $\theta > 0$ and symmetric positive definite (covariance) matrix $\Sigma \in \mathbb{R}^{d\times d}$ this symmetric random walk Metropolis algorithm is valid: $I_n \to I$ almost surely as $n \to \infty$ (e.g. [14, Theorem 1]). However, the efficiency of the method, that is, the speed at which I_n converges to I, is crucially affected by the choice of θ and Σ. Suppose for a moment that the matrix Σ is fixed, and we only vary $\theta > 0$. Then, for too large θ, few proposals become accepted and the chain mixes poorly. If θ is too small, most of the proposals Y_n become accepted, but the steps $X_n - X_{n-1}$ are small, preventing good mixing. In fact, previous results indicate that the acceptance probability is closely related with the efficiency of the algorithm. Commonly used 'rule of thumb' is that the acceptance probability α_n should be on the average about 0.234 even though this choice is not always optimal [7,16,17,22]. In practice, such a θ is usually found by several trial runs, which can be laborious and time-consuming.

So called adaptive MCMC algorithms have gained popularity since the seminal work of Haario et al. [11]. Several other such algorithms have been proposed after Andrieu and Robert [3] noticed the connection between Robbins–Monro stochastic approximation and adaptive MCMC [1,4,6,19,20]. The adaptive scaling Metropolis (ASM) algorithm optimises the scaling $\theta > 0$ of the proposal distribution adaptively, based on the observed acceptance probability. Namely, in step (S1) of the above algorithm, the constant θ is replaced, for example, with $\theta_{n-1} := e^{S_{n-1}}$ where $(S_n)_{n\geq 1}$ are random variables with $S_1 \equiv s_1 \in \mathbb{R}$ and for $n \geq 2$ defined recursively as follows

(S3) $S_n = S_{n-1} + \eta_n(\alpha_n - \alpha^*)$

where the parameter α^* determines the desired mean acceptance probability, often 0.234, and $(\eta_n)_{n\geq 2}$ is a sequence of positive adaptation step sizes decaying to zero.

A similar random walk Metropolis algorithm with adaptive scaling was actually proposed over a decade ago by Gilks et al. [10]. Their approach differed from the ASM approach so that the adaptation was performed only at particular regeneration times, which may occur infrequently or may be difficult to identify in practice. The ASM algorithm presented above has been proposed earlier by several authors [3,6,20], with a slightly different update formula (S3). The exact form of (S3) was used by Andrieu and Thoms [4] and Atchadé and Fort [5]. The crucial difference of the present paper compared to the earlier works is that the algorithm does not involve any additional constraints on θ_n. This difference is chiefly a theoretical advance, as discussed below. Therefore, no empirical studies of the performance of the algorithms are included in the paper.

Since the ASM algorithm only adapts the scale of the proposal distribution, it is likely to be inefficient in certain situations. For example, if π is high-dimensional and possesses a strong correlation structure and Σ does not match this structure, the ASM approach is likely to be suboptimal. In such a situation, one can employ the Adaptive Metropolis (AM) algorithm [11] to adapt the covariance shape with the scaling adaptation [4,5]. That is, in addition to using random θ_{n-1} in (S1), one also uses a random matrix Σ_{n-1} instead of a fixed Σ. Namely, Σ_n is a covariance estimator based on $X_1, \ldots, X_n$; the details can be found in Section 2. This algorithm will be referred here to as the adaptive scaling within AM (ASWAM).

It is not obvious that adaptive algorithms like the ASM and the ASWAM are valid, that is, $I_n \to I$. In fact, there are examples of continuously adapting MCMC schemes that destroy the correct ergodic[1] properties [19]. Many ergodicity results on adaptive MCMC algorithms in the literature assume some 'uniform' behaviour for all the possible MCMC kernels [5,6,19]. In the context of the adaptive scaling framework, this essentially means that θ_n must be constrained to a predefined set $[a, b]$ with some $0 < a \leq b < \infty$. Recent findings of Saksman and Vihola [21] allow one also to prove the ergodicity in a non-uniform case; Fort et al. [8] elaborate this approach in a much more general setting. In order to employ these results, one would typically enforce $\theta_n \in [a_n, b_n]$ where the sequences $a_n \searrow 0$ and $b_n \nearrow \infty$ with a certain rate. Alternatively, one can use a general reprojection technique due to Andrieu and Moulines [1] on the sets $[a_n, b_n]$ without the rate assumption on a_n and b_n but with the cost of possible 'restarts' of the process. It is also possible to modify the adaptation rule to ensure stable behaviour [4]. Such constraints and stabilisation structures are theoretically convenient, but may pose a problem for a practitioner. Good values for the constraint parameters may be difficult to choose without prior knowledge of the target distribution π. In the worst case, the values are chosen inappropriately and the algorithm is rendered useless in practice.

It is a common belief that many of the proposed adaptive MCMC algorithms are inherently stable and thereby do not require additional constraints or stabilisation structures. Indeed, there is considerable empirical evidence of the stability of several unconstrained algorithms, including the adaptive scaling approach. There are yet only few theoretical results, especially Saksman and Vihola [21] verifying the correct ergodic properties and the stability of the AM algorithm [11], provided the target distribution π has super-exponentially decaying tails with regular contours. These assumptions on π are close to those that ensure the geometric ergodicity of a non-adaptive random walk Metropolis algorithm [13]. The result in [21] does not assume an upper bound, but requires an explicit lower bound for the adapted covariance parameter.[2] In the context of the scaling adaptation, the lower bound is analogous to constraining θ_n to the interval $[a, \infty)$, where $a > 0$.

The main results of this paper, formulated in the next section, show that the stability and ergodicity of the ASM algorithm can be verified under similar assumptions on the target distribution as in [21], without any modifications or constraints on the adaptation parameter $\theta_n \in (0, \infty)$. These are the first results that validate the correctness of a completely unconstrained, fully adaptive MCMC algorithm. A similar result applies for the ASWAM approach, given that stability is enforced on the covariance parameter Σ_n by bounding the eigenvalues away from zero and infinity.

2. Main results

The scaling adaptation introduced in Section 1 can be generalised by considering a function ϕ mapping real-valued parameter values S_n to a scaling in $(0, \infty)$.

[1] In the present work, the word 'ergodicity' refers to the convergence of ergodic averages I_n to I, unlike Roberts and Rosenthal [19] who define 'ergodic' by the convergence of the marginal distributions of X_n to π in the total variation sense.

[2] The recent work [23] gives partial stability results of the AM also without the lower bound.

Assumption 1. The scaling function $\phi : \mathbb{R} \to (0, \infty)$ is increasing and surjective, piecewise differentiable and there are constants $h, c > 0$ and $\kappa \geq 1$ such that

$$\phi'(s + \bar{h}) \leq c \max\{1, \phi^{\kappa}(s)\}$$

for all $s \in \mathbb{R}$ and all $0 \leq \bar{h} \leq h$.

The function $\phi(s) = e^s$ was suggested above, but Assumption 1 allows one to use also, for example, piecewise polynomially defined ϕ. For example, defining $\phi(x) = x$ whenever x is greater than some $x_0 > 0$ and continuing ϕ appropriately for $x < x_0$ gives an algorithm in the spirit of Atchadé and Rosenthal [6].

The results hold also for other than a Gaussian proposal, as long as the proposal density is spherically symmetric and satisfies a certain tail behaviour.

Assumption 2. The proposal density q can be written as $q(z) = \hat{q}(\|z\|)$ where $\hat{q} : [0, \infty) \to (0, \infty)$ is a bounded, decreasing and differentiable function. Moreover, for any $\xi \in (0, 1)$ there exist an $\epsilon^* > 0$, constants $0 \leq a < b < \infty$ and $c_1, c_2, c_3 > 0$ such that for all $\epsilon \in [0, \epsilon^*]$, the following bounds hold for the derivative of $\hat{q}$

$$\xi \hat{q}'(x) - \hat{q}'(x + \epsilon) \geq c_1, \quad \text{for all } a \leq x \leq b,$$
$$\int_0^{\infty} \min\{0, \xi \hat{q}'(x) - \hat{q}'(x + \epsilon)\} \mathrm{d}x \geq -c_2 e^{-c_3 \epsilon^{-1}}.$$

Proposition 27 in Appendix B shows that Assumption 2 holds for Gaussian and Student distributions q.

We also need certain conditions for the adaptation step size sequence $(\eta_n)_{n \geq 2}$.

Assumption 3. The sequence $(\eta_n)_{n \geq 2}$ is non-negative, $\sum_{n=2}^{\infty} \eta_n = \infty$ and $\sum_{n=2}^{\infty} \eta_n^2 < \infty$.

Assumption 3 is classical in the context of stochastic approximation. A typical choice for the step size sequence satisfying Assumption 3 is $\eta_n \propto n^{-\gamma}$ with some constant $\gamma \in (1/2, 1]$.

We are now ready to define the adaptive scaling Metropolis (ASM) and the adaptive scaling within adaptive Metropolis (ASWAM) algorithms.

Definition 4 (*ASM*)**.** Suppose that the matrix $\Sigma \in \mathbb{R}^{d \times d}$ is symmetric and positive definite, ϕ satisfies Assumption 1, q satisfies Assumption 2 and $(\eta_n)_{n \geq 2}$ satisfies Assumption 3. Let $\{U_n, W_n\}_{n \geq 2}$ be a set of independent random variables where each U_n is uniformly distributed in the unit interval $[0, 1]$ and each W_n has the distribution q for all $n \geq 2$. Let $X_1 \equiv x_1 \in \mathbb{R}^d$ with $\pi(x_1) > 0$ and $S_1 \equiv s_1 \in \mathbb{R}$, and for $n \geq 2$ define recursively

$$Y_n = X_{n-1} + \phi(S_{n-1}) \Sigma^{1/2} W_n \tag{1}$$

$$X_n = \begin{cases} Y_n, & \text{if } U_n \leq \alpha_n \\ X_{n-1}, & \text{otherwise} \end{cases} \tag{2}$$

$$S_n = S_{n-1} + \eta_n (\alpha_n - \alpha^*), \tag{3}$$

where $\alpha_n := \min\{1, \pi(Y_n)/\pi(X_{n-1})\}$ stands for the acceptance probability.

Definition 5 (*ASWAM*)**.** Assume the setting of the ASM algorithm in 4, but instead of (1) use

$$Y_n = X_{n-1} + \phi(S_{n-1}) \Sigma_{n-1}^{1/2} W_n. \tag{4}$$

The covariance process $(\Sigma_n)_{n\geq 1}$ is determined as follows: let $\mu_1 \equiv x_1 \in \mathbb{R}^d$, suppose $\Sigma_1 \in \mathbb{R}^{d\times d}$ is a symmetric and positive definite matrix and

$$\hat{\mu}_n = (1-\eta_n)\mu_{n-1} + \eta_n X_n \tag{5}$$

$$\hat{\Sigma}_n = (1-\eta_n)\Sigma_{n-1} + \eta_n (X_n - \mu_{n-1})(X_n - \mu_{n-1})^T \tag{6}$$

$$(\mu_n, \Sigma_n) = \begin{cases} (\hat{\mu}_n, \hat{\Sigma}_n), & \text{if } (\hat{\mu}_n, \hat{\Sigma}_n) \in \mathbb{S}_\zeta \text{ and} \\ (\mu_{n-1}, \Sigma_{n-1}), & \text{otherwise,} \end{cases} \tag{7}$$

where the truncation set is defined as $\mathbb{S}_\zeta = \{(\mu, \Sigma) : \|\mu\| \leq \zeta,\ \lambda(\Sigma) \subset [\zeta^{-1}, \zeta]\}$ with $\lambda(\Sigma)$ being the set of the eigenvalues of Σ and $\zeta \in [1, \infty)$ is a constant parameter.

The step (7) enforces the stability of the covariance adaptation process, while the scaling parameter S_n follows Eq. (3).

Before stating the first ergodicity result, consider the following condition on the regularity of a collection of sets. Before that, recall that a C^1 domain in $\mathbb{R}^d$ is a domain whose boundary is locally a graph of a continuously differentiable function.

Definition 6. Suppose that $\{A_i\}_{i\in I}$ is a collection of sets $A_i \subset \mathbb{R}^d$ each consisting of finitely many disjoint components that are closures of C^1 domains. Let $n_i(x)$ stand for the outer-pointing normal at x in the boundary ∂A_i. Then, $\{A_i\}_{i\in I}$ have *uniformly continuous normals* if for all $\epsilon > 0$ there is a $\delta > 0$ such that for any $i \in I$ it holds that $\|n_i(x) - n_i(y)\| \leq \epsilon$ for all $x, y \in \partial A_i$ such that $\|x - y\| \leq \delta$.

Definition 6 essentially states that the boundaries ∂A_i must be regular enough to ensure that if one looks at any ∂A_i at a sufficiently small scale, it will look locally almost like a plane.

Theorem 7. *Assume π has a compact support $\mathbb{X} \subset \mathbb{R}^d$ and π is continuous and bounded away from zero on $\mathbb{X}$. Moreover, assume that $\mathbb{X}$ has a uniformly continuous normal (Definition 6) and $\alpha^* \in (0, \frac{1}{2})$. Then, for either the ASM or the ASWAM process and for any bounded function f, the strong law of large numbers holds that is,*

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x \quad \textit{almost surely.} \tag{8}$$

The proof of Theorem 7 is given in Section 5.

Let us consider next target distributions π with unbounded supports, satisfying the following conditions formulated in [21].

Assumption 8. The density π is bounded, bounded away from zero on compact sets, differentiable, and

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|^\rho} \cdot \nabla \log \pi(x) = -\infty \tag{9}$$

for some constant $\rho > 1$, where $\|\cdot\|$ stands for the Euclidean norm. Moreover, the contour normals satisfy

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|} \cdot \frac{\nabla\pi(x)}{\|\nabla\pi(x)\|} < 0. \tag{10}$$

This assumption is very near to the conditions introduced by Jarner and Hansen [13] to ensure the geometric ergodicity of a (non-adaptive) Metropolis algorithm, and considered by Andrieu and Moulines [1] in the context of adaptive MCMC. In particular, [1,13] assume that π fulfils the contour regularity condition (10). Instead of (9), they assume a super-exponential decay on π,

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

which is only slightly more general than (9) allowing $\rho = 1$. See [13] for examples and discussion on these conditions.

Theorem 9. *Suppose $\alpha^* \in (0, \frac{1}{2})$, π fulfils Assumption 8 and there is a $t_0 > 0$ such that the collection of contour sets $\{x \in \mathbb{R}^d : \pi(x) \geq t\}_{0<t\leq t_0}$ have uniformly continuous normals (Definition 6). Assume that there exist constants $c < \infty$ and $p \in (0, 1)$ such that $|f(x)| \leq c\pi^{-p}(x)$ for all $x \in \mathbb{R}^d$. Then, for the ASM and the ASWAM processes, the strong law of large numbers (8) holds.*

The proof of Theorem 9 is given in Section 5.

Remark 10. For many practical target densities satisfying Assumption 8 the tail contours are (essentially) scaled copies of each other, in which case they have automatically uniformly continuous normals. This indicates that the conditions of Theorem 9 are practically similar to [21, Theorem 10] verifying the ergodicity of the Adaptive Metropolis algorithm.

Remark 11. The 'safe' values for the desired acceptance rate stipulated by Theorems 7 and 9 are $\alpha^* \in (0, 1/2)$. The values $[1/2, 1)$ are excluded due to technical reasons, in particular due to Proposition 17 establishing the lower bound for $\phi(S_n)$. It is expected that Theorems 7 and 9 hold assuming only $\alpha^* \in (0, 1)$, but this cannot be verified with the present approach. The range $\alpha^* \in (0, 1/2)$ is, however, often sufficient in practice, as the most commonly used values for a random walk Metropolis algorithm are probably $\alpha^* = 0.234$ and $\alpha^* = 0.44$, and it has been suggested that values $\alpha^* \in [0.1, 0.4]$ should work well in most cases [7,16,17,20].

Remark 12. The conditions on the proposal density in Assumption 2 are not optimal. The technical tail decay condition on $\hat{q}$ is needed in the case of π with an unbounded support in Theorem 9. Theorem 7 considering compactly supported π can be established for a more general class of proposal distributions, but this is not pursued here.

Remark 13. Theorems 7 and 9 ensure that the trajectories of the ergodic averages converge almost surely but do not state explicit results on the convergence of the marginal distributions of X_n. The marginal convergence (in the total variation sense) could be established using the stability results in Section 4 and the recent work of Fort et al. [8].

The rest of the article is organised as follows. Section 3 describes a general framework for scale adaptation covering simultaneously both the ASM and the ASWAM algorithms. Section 4 develops stability results for this process. In particular, Corollary 19 ensures the stability of the sequence $\phi(S_n)$ with the assumptions of Theorem 7, and Proposition 20 controls the growth of $\phi(S_n)$ when π fulfils the conditions of Theorem 9. Once the stability results are obtained, Theorems 7 and 9 are proved in Section 5 using the results in [21].

3. Framework and notation

Consider a process $(X_n, \Gamma_n)_{n\geq 1}$ evolving in the measurable space $\mathbb{X}\times\mathbb{G}$, where the support of the target density $\mathbb{X} := \{x \in \mathbb{R}^d : \pi(x) > 0\}$ is the space of the 'MCMC' chain $(X_n)_{n\geq 1}$, and the adaptation parameters $(\Gamma_n)_{n\geq 1} = (S_n, \mu_n, \Sigma_n)_{n\geq 1}$ evolve in $\mathbb{G} = \mathbb{R} \times \mathbb{S}_\zeta$; the scaling parameters $(S_n)_{n\geq 1}$ are real-valued and the covariance adaptation process $(\mu_n, \Sigma_n)_{n\geq 1}$ takes values on the space $\mathbb{S}_\zeta \subset \mathbb{R}^d \times \mathcal{C}_\zeta$ with

$$\mathcal{C}_\zeta := \{\Sigma \in \mathbb{R}^{d\times d} : \Sigma \text{ is symmetric and } \lambda(\Sigma) \subset [\zeta^{-1}, \zeta]\}$$

and where $\lambda(\Sigma)$ stands for the set of eigenvalues of Σ. By this definition, we may define $\mathbb{S}_\zeta = \{(\mu, \Sigma)\}$ in the case of the ASM whence $\Sigma_n = \Sigma$ and $\mu_n = \mu$ for all $n \geq 1$ and for the ASWAM, (μ_n, Σ_n) is determined through (5)–(7). We need the specific form of adaptation of (μ_n, Σ_n) only in Section 5. For the stability results in Section 4 it is sufficient that $\Sigma_n \in \mathcal{C}_\zeta$.

Denote $\mathcal{F}_n := \sigma(W_n, U_n : 1 \leq k \leq n)$ so that $(\mathcal{F}_n)_{n\geq 1}$ is a filtration and also each Γ_n is $\mathcal{F}_n$-adapted. With these definitions, we may write

$$Y_{n+1} \mid \mathcal{F}_n \sim q_{\Gamma_n}(X_n, \cdot) \tag{11}$$

$$X_{n+1} = Y_{n+1}\mathbb{1}_{\{U_{n+1}\leq \alpha_{n+1}\}} + X_n \mathbb{1}_{\{U_{n+1}>\alpha_{n+1}\}} \tag{12}$$

$$S_{n+1} = S_n + \eta_{n+1} H(X_n, Y_{n+1}) \tag{13}$$

where $\mathbb{1}_A$ stands for the indicator function of a set A and $H(x, y) := \alpha(x, y) - \alpha^*$ with $\alpha(x, y) := \min\{1, \frac{\pi(y)}{\pi(x)}\}$. Moreover, for $\gamma = (s, \mu, \Sigma) \in \mathbb{G}$ the proposal density is defined as

$$q_\gamma(z) = q_{(s,\Sigma)}(z) = [\phi(s)]^{-d} \det(\Sigma)^{-1/2} q([\phi(s)]^{-1}\Sigma^{-1/2} z). \tag{14}$$

Note that the form (13) of adaptation can be considered as the Robbins–Monro stochastic approximation; see [1–3] and references therein.

We will need the notion of expected acceptance rate at $x \in \mathbb{X}$ with parameter $\gamma \in \mathbb{G}$ as

$$\mathrm{acc}(x, \gamma) := \int_{\mathbb{X}} \alpha(x, y) q_\gamma(x - y)\mathrm{d}y.$$

On average, the adaptation rule decreases S_n whenever $\mathrm{acc}(X_n, \Gamma_n) < \alpha^*$, and vice versa. So, it is plausible to expect that the algorithm would eventually result in $\Gamma_n \to \gamma^* \in \mathbb{G}$ such that the overall expected acceptance rate $\int_{\mathbb{X}} \mathrm{acc}(x, \gamma)\pi(x)\mathrm{d}x = \alpha^*$. In this paper, however, the convergence of Γ_n is not the main concern, but the stability of it, as it turns out to be crucial for the validity of the algorithms considered.

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

$$P_\gamma(x, A) := \mathbb{1}_A(x) \int_{\mathbb{R}^d} [1 - \alpha(x, y)] q_\gamma(x - y)\mathrm{d}y + \int_A \alpha(x, y) q_\gamma(x - y)\mathrm{d}y. \tag{15}$$

Using the kernels P_γ, one can write (11) and (12) as $\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = P_{\Gamma_n}(X_n, A)$. As usual, integration of a function f with respect to a transition kernel is denoted as

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

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

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

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

4. Stability

This section develops stability results for the general adaptive scaling process of Section 3. We start with a general stability theorem based on a martingale argument. This theorem is auxiliary for the present paper, but may have applications also in other settings.

Theorem 14. *Suppose $(\mathcal{F}_n)_{n\ge1}$ is a filtration, $(\eta_n)_{n\ge2}$ are non-negative constants such that $\sum \eta_n^2 < \infty$ and H_n are $\mathcal{F}_n$-adapted random variables satisfying* $\limsup_{n\to\infty} \eta_n H_n \le 0$ *and*

$$\sum_{n=2}^{\infty} \eta_n^2 (\mathbb{E}[H_n^2 \mid \mathcal{F}_{n-1}] - \mathbb{E}[H_n \mid \mathcal{F}_{n-1}]^2) < \infty. \tag{16}$$

Let $S_1 \equiv s_1 \in \mathbb{R}$, and define $S_{n+1} := S_n + \eta_{n+1} H_{n+1}$ recursively for all $n \ge 1$.

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

$$\mathbb{E}[H_{n+1} \mathbb{1}_{\{S_n \ge a\}} \mid \mathcal{F}_n] \le 0,$$

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

(ii) *If also $\sum \eta_n = \infty$ and there is a non-decreasing sequence of $\mathcal{F}_n$-adapted random variables $(A_n)_{n\ge1} \subset \mathbb{R}$ and a constant $b < 0$ such that for all $n \ge 1$*

$$\mathbb{E}[H_{n+1} \mathbb{1}_{\{S_n \ge A_n\}} \mid \mathcal{F}_n] \le b \mathbb{1}_{\{S_n \ge A_n\}},$$

then $\limsup_{n\to\infty}(S_n - A_n) \le 0$ *a.s.*

Proof. Let $W_n := H_n \mathbb{1}_{\{S_{n-1} \ge a\}}$ for $n \ge 2$, and define the martingale $(M_n, \mathcal{F}_n)_{n\ge1}$ by setting $M_1 := 0$, and $M_n := \sum_{k=2}^n \mathrm{d}M_k$ for $n \ge 2$ with the differences $\mathrm{d}M_n := \eta_n (W_n - \mathbb{E}[W_n \mid \mathcal{F}_{n-1}])$. Now,

$$\sum_{k=2}^{\infty} \mathbb{E}[\mathrm{d}M_k^2 \mid \mathcal{F}_{k-1}] = \sum_{k=2}^{\infty} \eta_k^2 (\mathbb{E}[H_n^2 \mid \mathcal{F}_{n-1}] - \mathbb{E}[H_n \mid \mathcal{F}_{n-1}]^2) \mathbb{1}_{\{S_{n-1} \ge a\}} < \infty$$

by assumption. This implies that almost every path of M_n converges to a finite limit M_∞ (e.g. [12, Theorem 2.15]).

Let $(\tau_k)_{k\ge1}$ be the exit times of S_n from $(-\infty, a)$, defined as $\tau_k := \inf\{n > \tau_{k-1} : S_n \ge a,\ S_{n-1} < a\}$ using the conventions $\tau_0 = 0$, $S_0 < a$, and $\inf \emptyset = \infty$. Define also the latest exit from $(-\infty, a)$ until time n by $\sigma_n := \sup\{\tau_k : k \ge 1,\ \tau_k \le n\}$. Whenever $S_n \ge a$, one can write $S_n = S_{\sigma_n} + (M_n - M_{\sigma_n}) + Z_{\sigma_n, n}$ where

$$Z_{m,n} := \sum_{k=m+1}^{n} \eta_k \mathbb{E}[W_k \mid \mathcal{F}_{k-1}] \le 0$$

by assumption. In this case,

$$\begin{aligned} S_n &\le S_{\sigma_n} + (M_n - M_{\sigma_n}) \le \max\{S_1, a\} + \eta_{\sigma_n} H_{\sigma_n} + |M_n| + |M_{\sigma_n}| \\ &\le \max\{S_1, a\} + \sup_{k\ge1} \eta_k H_k + 2 \sup_{k\ge1} |M_k| \le C \end{aligned} \tag{17}$$

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

Assume then (ii). If $S_n < A_n$ for all n greater than some $N_1(\omega) < \infty$, the claim is trivial. Suppose then that $S_n \geq A_n$ infinitely often. Define $(\tau_k)_{k\geq 1}$ as the exit times of S_n from $(-\infty, A_n)$ as above, $\tau_k := \inf\{n > \tau_{k-1} : S_n \geq A_n,\ S_{n-1} < A_{n-1}\}$ with $\tau_0 \equiv 0$ and $S_0 < A_0$. The times τ_k must be a.s. finite in this case (and S_n returns to $(-\infty, A_n)$ infinitely often), for suppose the contrary: then the last exit times σ_n are bounded by some $\sigma_n \leq \sigma < \infty$, and for $n \geq \sigma$ one may write

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

where M_n and $Z_{n,m}$ are defined as above, but using the random variables $W_n := H_n \mathbb{1}_{\{S_{n-1}\geq A_{n-1}\}}$, and the random variable C_σ is a.s. finite as in (17). Now, $Z_{\sigma,n} \to -\infty$ a.s. as $n \to \infty$, so $S_n < A_n$ a.s. for sufficiently large n.

Consider then the case $(\tau_k)_{k\geq 1}$ are all finite and M_n converges to a finite M_∞. Fix an $\epsilon > 0$ and let $N_0 = N_0(\omega, \epsilon)$ be such that for all $n \geq N_0$, it holds that $\eta_{\sigma_n} H_{\sigma_n} \leq \epsilon/3$ and that $|M_k - M_\infty| \leq \epsilon/3$ a.s. for all $k \geq \sigma_n$. The claim follows from the estimate

$$\begin{aligned} S_n &\leq S_{\sigma_n} + (M_n - M_{\sigma_n}) = S_{\sigma_n - 1} + \eta_{\sigma_n} H_{\sigma_n} + (M_n - M_{\sigma_n}) \\ &\leq A_{\sigma_n} + \epsilon/3 + |M_n - M_\infty| + |M_\infty - M_{\sigma_n}| \leq A_n + \epsilon \end{aligned}$$

for all $n \geq N_0$. $\square$

Hereafter, we shall consider the adaptive scaling process described in Section 3. One can give simple conditions under which the result of Theorem 14 applies, since

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

so by the boundedness of H it is sufficient to find out when $\mathrm{acc}(x, \gamma)$ is below or above α^*.

Lemma 15. *Suppose q satisfies Assumption 2 and $q_{(s,\Sigma)}$ is defined through (14). Then, there exists a constant $\bar{c} < \infty$ such that*

$$\sup_{z\in\mathbb{R}^d,\ \Sigma\in\mathcal{C}_\zeta} q_{(s,\Sigma)}(z) \leq \bar{c}[\phi(s)]^{-d} \quad \text{for all } s \in \mathbb{R}. \tag{18}$$

Moreover, for any $\epsilon > 0$ there exist $M < \infty$ such that for all $s \in \mathbb{R}$ and any plane $P \subset \mathbb{R}^d$

$$\inf_{\Sigma\in\mathcal{C}_\zeta} \int_{\overline{B}(0,\phi(s)M)} q_{(s,\Sigma)}(z)\mathrm{d}z \geq 1 - \epsilon \tag{19}$$

$$\sup_{\Sigma\in\mathcal{C}_\zeta} \int_{\{d(z,P)\leq\phi(s)M^{-1}\}} q_{(s,\Sigma)}(z)\mathrm{d}z \leq \epsilon. \tag{20}$$

The proof of Lemma 15 is straightforward; the details are given in Appendix A.

Let us then record a simple estimate on the expected acceptance rate when π is compact and S_n is large.

Proposition 16. *Suppose q satisfies Assumption 2 and π is supported on a compact set $\mathbb{X} \subset \mathbb{R}^d$ and $\alpha^* > 0$. Then, there is $b < 0$ and $a \in \mathbb{R}$ such that*

$$\mathbb{E}[H(X_n, Y_{n+1}) \mid \mathcal{F}_n] \leq b \quad \text{whenever } S_n \geq a. \tag{21}$$

Proof. Compute for any $x \in \mathbb{X}$ and all $\gamma = (s, \mu, \Sigma) \in \mathbb{G}$

$$\begin{aligned}\mathrm{acc}(x,\gamma) &= \int_{\mathbb{R}^d} \alpha(x,y) q_\gamma(x-y)\mathrm{d}y \le \int_{\overline{B}(x,\mathrm{diam}(\mathbb{X}))} q_\gamma(z)\mathrm{d}z \\ &\le \int_{\overline{B}(x,\mathrm{diam}(\mathbb{X}))} \sup_{\Sigma\in\mathcal{C}_\zeta} q_{(s,\Sigma)}(z)\mathrm{d}z \le \bar{c}[\phi(s)]^{-d} \int_{\overline{B}(0,\mathrm{diam}(\mathbb{X}))} \mathrm{d}z\end{aligned}$$

by (18) in Lemma 15. We may choose a to be sufficiently large so that $\mathrm{acc}(x,\gamma) \le \alpha^*/2$ whenever $s \ge a$. That is, (21) holds with $b = -\alpha^*/2 < 0$, whenever $S_n \ge a$. □

Next, we shall consider the case S_n small, simultaneously for both cases where π is compactly supported and π has a super-exponential tail.

Proposition 17. *Suppose that there is a $t_0 > 0$ such that $L_{t_0} := \{y \in \mathbb{R}^d : \pi(y) \ge t_0\}$ is compact and π is continuous on L_{t_0}. Moreover, suppose that the sets in the collection $\{L_t\}_{0<t\le t_0}$ have uniformly continuous normals (Definition 6) and q satisfies Assumption 2. Then, for any $\alpha^* < 1/2$, there are $a \in \mathbb{R}$ and $b > 0$ such that*

$$\mathbb{E}[H(X_n, Y_{n+1}) \mid \mathcal{F}_n] \ge b \quad \textit{whenever } S_n \le a. \tag{22}$$

Before giving the proof of Proposition 17, let us outline the simple intuition behind it. For all s small enough and for any $\Sigma \in \mathcal{C}_\zeta$, the mass of $q_{(s,\Sigma)}$ is essentially concentrated on a small ball $\overline{B}(0,\epsilon)$. If one looks the target π only on $\overline{B}(x,\epsilon)$, there are, roughly speaking, two alternatives. The first one is that π is approximately constant on that small ball and $\mathrm{acc}(x,\gamma) \approx 1$. The second alternative is that π decreases very rapidly to *one* direction, in which case the set $\{y : \pi(y) \ge \pi(x)\}$ looks like a half-space on the ball $\overline{B}(x,\epsilon)$, and consequently $\mathrm{acc}(x,\gamma) \gtrsim 1/2$.

Before the proof, we shall formulate a lemma on this 'half-space approximation.'

Lemma 18. *Suppose that the sets $\{A_i\}_{i\in I}$ with $A_i \subset \mathbb{R}^d$ have uniformly continuous normals (Definition 6). Then, for any $\epsilon > 0$, there is a $\delta > 0$ such that for any $i \in I$, any $x \in A_i$ and any $r \in (0,\delta]$, there is a half-space T such that $\overline{B}(x,r) \cap T \subset \overline{B}(x,r) \cap A_i$, and the distance $d(x,T) \le \epsilon r$.*

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

Proof of Proposition 17. Fix an $\epsilon^* \in (0,1)$ and let $M = M(\epsilon^*)$ be the constant from Lemma 15 applied with $\epsilon = \epsilon^*$.

By compactness of L_{t_0} and continuity of π one can find $\delta_1 > 0$ such that for all $x, y \in L_{t_0}$ with $\|x-y\| \le \delta_1$, it holds that $|\log\pi(x) - \log\pi(y)| \le \epsilon^*$ so that

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

Let $\delta_2 > 0$ be sufficiently small to satisfy Lemma 18 with the choice $\epsilon = M^{-2}$.

Choose a small enough $a \in \mathbb{R}$ so that $2\phi(a)M \le \min\{\delta_1, \delta_2\}$. Let $s \le a$, denote $r_s := \phi(s)M$, and write for any $x \in L_{t_0}$

$$\begin{aligned}\int_{\mathbb{X}} \alpha(x,y) q_\gamma(x-y)\mathrm{d}y &\ge \int_{\overline{B}(x,r_s)\cap L_{t_0}} \alpha(x,y) q_\gamma(x-y)\mathrm{d}y \\ &\ge (1-\epsilon^*) \int_{\overline{B}(x,r_s)\cap L_{t_0}} q_\gamma(x-y)\mathrm{d}y\end{aligned}$$

since $2r_s \leq \delta_1$. Denote by T the half-space from Lemma 18, such that $\overline{B}(x, r_s) \cap T \subset \overline{B}(x, r_s) \cap L_{t_0}$ and the distance $d(x, T) \leq M^{-2} r_s$. One obtains

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

where $\tilde{T}$ is the half-space with the boundary plane P parallel to the boundary of T, and passing through x. Lemma 15 yields the last inequality, specifically (19) with the symmetry of q_γ and (20). The same estimate clearly holds for any $x \in L_t$ with $t \in (0, t_0)$.

To conclude, for any $\alpha^* < 1/2$ one can choose a sufficiently small $\epsilon^* = \epsilon^*(\alpha^*) > 0$ such that for all $x \in \mathbb{X}$ and for any $\gamma = (s, \mu, \Sigma)$ with $s \leq a$

$$\mathrm{acc}(x, \gamma) = \int_{\mathbb{X}} \alpha(x, y) q_\gamma(x-y)\mathrm{d}y \geq \frac{1}{2} - \frac{1}{2}\left(\frac{1}{2} - \alpha^*\right).$$

This implies (22) with $b = (1/2 - \alpha^*)/2 > 0$. □

As an easy corollary of the propositions above, one establishes the stability of the adaptive scaling process on the case of compactly supported π.

Corollary 19. *Suppose q and $(\eta_n)_{n\geq 2}$ satisfy Assumptions 2 and 3, respectively, π has a compact support $\mathbb{X} \subset \mathbb{R}^d$ and π is continuous, bounded and bounded away from zero on $\mathbb{X}$. Moreover, assume that $\mathbb{X}$ has a uniformly continuous normal (Definition 6). Then, for the general adaptive scaling process in Section 3 with any $\alpha^* \in (0, \frac{1}{2})$ there exist a.s. finite random variables A_1 and A_2 such that for all $n \geq 1$*

$$A_1 \leq S_n \leq A_2. \tag{23}$$

Proof. The conditions of Propositions 16 and 17 are satisfied, so there are constants $-\infty < a_1 < a_2 < \infty$ and $b < 0$ such that

$$\mathbb{E}[H(X_n, Y_{n+1}) \mid \mathcal{F}_n] \leq b \quad \text{whenever } S_n \geq a_2,$$
$$\mathbb{E}[H(X_n, Y_{n+1}) \mid \mathcal{F}_n] \geq -b \quad \text{whenever } S_n \leq a_1.$$

Theorem 14 can be applied to $-S_n$ and S_n, since by the boundedness of H (16) is implied by $\sum \eta_n^2 < \infty$. Theorem 14 guarantees that $a_1 \leq \liminf_{n\to\infty} S_n$ and $\limsup_{n\to\infty} S_n \leq a_2$, respectively, from which one obtains a.s. finite A_1 and A_2 for which (23) holds. □

The rest of this section considers targets π with an unbounded support. Under a suitably regular π, it is shown that the growth of S_n can be controlled. The following estimate for the at most polynomial growth of $\phi(S_n)$ is crucial for the ergodicity result in Theorem 9.

Proposition 20. *Suppose π fulfils Assumption 8 and there is a $t_0 > 0$ such that the collection of contour sets $\{x \in \mathbb{R}^d : \pi(x) \geq t\}_{0<t\leq t_0}$ have uniformly continuous normals (Definition 6). Suppose also that ϕ, q and $(\eta_n)_{n\geq 2}$ satisfy Assumptions 1, 2 and 3, respectively. Then, for the*

general adaptive scaling process in Section 3 with $\alpha^* \in (0, \frac{1}{2})$*, and for any* $\beta > 0$*, there exist an a.s. positive* $\Theta_1 = \Theta_1(\omega)$ *and an a.s. finite* $\Theta_2 = \Theta_2(\omega, \beta)$ *such that for all* $n \geq 1$

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

Before the proof, let us consider an estimate of $\mathrm{acc}(x, (s, \mu, \Sigma))$ depending on both x and s.

Lemma 21. *Assume* q *satisfies Assumption 2 and* π *satisfies Assumption 8. Then, for any* $\epsilon > 0$*, there is a constant* $c = c(\epsilon) \geq 1$ *such that* $\mathrm{acc}(x, (s, \mu, \Sigma)) \leq \epsilon$ *for all* $\phi(s) \geq c \max\{1, \|x\|\}$.

Proof. Let $r_1 \geq 1$ be sufficiently large so that for some $\nu > 0$ it holds that $\frac{x}{\|x\|} \cdot \frac{\nabla\pi(x)}{\|\nabla\pi(x)\|} < -\nu$ and $\frac{x}{\|x\|^\rho} \cdot \nabla \log \pi(x) < -\nu$ for all $\|x\| \geq r_1$. Increase r_1, if necessary, so that for any $\|x\| \geq r_1$ one can write $L_{\pi(x)} = \{y : \pi(y) \geq \pi(x)\} = \{ru : u \in S^d, 0 \leq r \leq g(u)\}$ where $S^d := \{u \in \mathbb{R}^d : \|u\| = 1\}$ is the unit sphere and the function $g : S^d \to (0, \infty)$ parametrises the boundary of $L_{\pi(x)}$. Notice also that the contour normal condition implies the existence of an $M \geq 1$ such that $L_{\pi(x)} \subset \overline{B}(0, M\|x\|)$ for all $\|x\| \geq r_1$ (see [21, Lemma 22]).

Write for $\|x\| \geq r_2 := Mr_1$ and denoting $T_x := \{d(y, L_{\pi(x)}) > \|x\|\}$

$$\begin{aligned}\mathrm{acc}(x, \gamma) &= \int_{\mathbb{R}^d} \alpha(x, y) q_\gamma(x - y)\mathrm{d}y \\ &\leq \int_{\mathbb{R}^d \setminus T_x} q_\gamma(x - y)\mathrm{d}y + \sup_{y \in \mathbb{R}^d} q_\gamma(x - y) \int_{T_x} \alpha(x, y)\mathrm{d}y.\end{aligned}$$

The first term can be estimated from above by (18) of Lemma 15

$$\int_{\overline{B}(0, M\|x\| + \|x\|)} q_\gamma(x - y)\mathrm{d}y \leq \bar{c}[\phi(s)]^{-d} \int_{\overline{B}(0, (M+1)\|x\|)} \mathrm{d}z \leq c_1[\phi(s)]^{-d}\|x\|^d \leq \frac{\epsilon}{2}$$

whenever $\phi(s) \geq (c_1 2/\epsilon)^{1/d}\|x\|$.

For the integral in the latter term, we use polar integration to estimate

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

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

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

from which

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

Consequently,

$$\int_{T_x} \alpha(x, y)\mathrm{d}y \leq c_d \frac{2}{\nu} \sup_{\tilde{g} \geq 1, \tilde{r} > 1} \exp\left[(d-1)\log(\tilde{g} + \tilde{r}) - \frac{\nu}{2}\tilde{g}^{\rho-1}\tilde{r}\right] \leq c_2$$

with a finite constant c_2 whenever $\|x\| \geq r_2$.

To sum up, there is a $c_3 > 0$ such that for any $\|x\| \geq r_2$ and any s satisfying

$$\phi(s) \geq c_3 \max\{1, \|x\|\} \geq \max\left\{\left(\frac{2c_1}{\epsilon}\right)^{1/d}\|x\|, \left(\frac{2\bar{c}c_2}{\epsilon}\right)^{1/d}\right\},$$

it holds that $\mathrm{acc}(x, (s, \mu, \Sigma)) \leq \epsilon$. For any $\|x\| < r_2$ there is a $r_2 \leq \|x_0\| \leq Mr_2$ such that $\pi(x_0) \leq \pi(x)$. Consequently, $\alpha(x, y) \leq \alpha(x_0, y)$ for all $y \in \mathbb{R}^d$ and therefore

$$\begin{aligned}\mathrm{acc}(x, \gamma) &\leq \int_{\mathbb{R}^d} \alpha(x_0, y) q_\gamma(x - y)\mathrm{d}y \\ &\leq \int_{\mathbb{R}^d \setminus T_{x_0}} q_\gamma(x - y)\mathrm{d}y + \sup_{y \in \mathbb{R}^d} q_\gamma(x - y) \int_{T_{x_0}} \alpha(x_0, y)\mathrm{d}y.\end{aligned}$$

Repeating the arguments above, there is a finite constant c_4 such that $\mathrm{acc}(x, (s, \mu, \Sigma)) \leq \epsilon$ for all $(\mu, \Sigma) \in \mathbb{S}_\zeta$ and for all $s \in \mathbb{R}$ such that $\phi(s) \geq c_4 \max\{1, \|x\|\}$. □

Having Lemma 21 and the lower bound from Proposition 17, the proof of Proposition 20 can be obtained by applying the growth condition on $\|X_n\|$ established in [21].

Proof of Proposition 20. Proposition 17 applied with Theorem 14 for $-S_n$ gives an a.s. finite A_1 such that $A_1 \leq S_n$ for all $n \geq 1$. The random variable $\Theta_1 := \phi(A_1)$ is a.s. positive, showing the lower bound.

To check the polynomial growth condition for $\phi(S_n)$, it is first verified that $\|X_n\|$ grows at most polynomially. Fix an $\epsilon > 0$ and let $\theta_1 = \theta_1(\epsilon) > 0$ and $a_1 = a_1(\epsilon) \in \mathbb{R}$ be such that $\theta_1 = \phi(a_1)$, and that $\mathbb{P}(B_1) \geq 1 - \epsilon$, with $B_1 := \{\Theta_1 \geq \theta_1\} = \{A_1 \geq a_1\}$. Let $V(x) := c_\pi \pi^{-1/2}(x)$, where the constant $c_\pi := [\sup_x \pi(x)]^{1/2}$ ensures that $V \geq 1$. Proposition 25 in Appendix B shows that the drift inequality

$$P_{(s,\Sigma)} V(x) \leq V(x) + b \tag{24}$$

holds for all $\Sigma \in \mathcal{C}_d$ and $\phi(s) \geq \theta_1 > 0$ with some $b = b(\theta_1) < \infty$. Construct an auxiliary process $(X'_n, \Gamma'_n)_{n \geq 1}$ coinciding with $(X_n, \Gamma_n)_{n \geq 1}$ in B_1 by setting $(X'_n, \Gamma'_n) = (X_{\tau_n}, \Gamma_{\tau_n})$ where the stopping times τ_n are defined as

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

Having the inequality (24), set $\beta' = \kappa^{-1}\beta$ where the constant $\kappa \geq 1$ is from Assumption 1 and use Proposition 7 of [21] to obtain the bound $\|X'_n\| \leq \Theta_\epsilon n^{\beta'}$ for some a.s. finite Θ_ϵ. The $\epsilon > 0$ was arbitrary, so one can let $\epsilon \to 0$ and obtain an a.s. finite Θ such that $\|X_n\| \leq \Theta n^{\beta'}$. Applying Lemma 21, one obtains that $\mathrm{acc}(X_n, (S_n, \Sigma_n)) \leq \alpha^*/2$ whenever $\phi(S_n) \geq \Theta' n^{\beta'}$ with $\Theta' := c_1 \max\{1, \Theta\}$.

Fix again an $\epsilon > 0$ and let $\theta_2 = \theta_2(\epsilon) < \infty$ be such that $\mathbb{P}(B_2) \geq 1 - \epsilon$ where $B_2 := \{\Theta' \leq \theta_2\}$. Construct an auxiliary process $(X'_n, S'_n)_{n \geq 1}$ coinciding with $(X_n, S_n)_{n \geq 1}$ in B_2 by stopping the process if $\phi(S_k) > \theta_2 k^{\beta'}$ as in the construction above. Theorem 14 ensures that

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

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

For $n \geq N_0$, one has

$$\phi(S'_n) \leq \phi(\tilde{a}_n + E_n) \leq c_2 \max\{1, \phi(\tilde{a}_n)^\kappa\} = c_2 \max\{1, \theta_2^\kappa n^{\kappa\beta'}\} \leq \theta'_2 n^\beta$$

for some finite θ'_2. Summing up, there is an a.s. finite Θ'_2 such that

$$\phi(S'_n) \leq \Theta'_2 n^\beta$$

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

5. Ergodicity

Section 4 established stability or controlled growth for the adaptive scaling process of Section 3. This section employs these results to prove strong laws of large numbers in Theorems 7 and 9 for the ASM and the ASWAM processes defined in Section 2, relying on the results introduced in [21]. For this purpose, consider the following theoretical adaptation framework introduced in [21] using a sequence of restriction sets $K_1 \subset K_2 \subset \cdots \subset K_n \subset \cdots \subset \mathbb{G}$.

Assume $(\tilde{X}_n, \tilde{Y}_n, \tilde{\Gamma}_n)_{n\geq 1}$ follow the general adaptation framework as described in Section 3. Assume $\tilde{\Gamma}_1 \equiv \tilde{\gamma}_1 \in K_1$ and instead of (13) let $(\tilde{\Gamma}_n)_{n\geq 1}$ follow the 'truncated' recursion

$$\tilde{\Gamma}_{n+1} = \sigma_{K_{n+1}}(\tilde{\Gamma}_n,\ \eta_{n+1}\hat{H}(\tilde{X}_n, \tilde{Y}_{n+1})) \tag{25}$$

where the restriction function $\sigma_K : \mathbb{G} \times \bar{\mathbb{G}} \to \mathbb{G}$ is defined as

$$\sigma_K(\gamma, \gamma') := \begin{cases} \gamma + \gamma', & \text{if } \gamma + \gamma' \in K \\ \gamma, & \text{otherwise,} \end{cases}$$

$\bar{\mathbb{G}} := \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d\times d} \supset \mathbb{G}$ and the function $\hat{H} : \mathbb{G} \times \mathbb{X}^2 \to \bar{\mathbb{G}}$ is defined as

$$\hat{H}((s, \mu, \Sigma), x, y) = \begin{bmatrix} H(x, y) \\ x - \mu \\ (x-\mu)(x-\mu)^T - \Sigma \end{bmatrix}.$$

That is, σ_{K_n} ensures that $\tilde{\Gamma}_n \in K_n$ for all $n \geq 1$. Observe that such a 'truncated process' can be constructed using an 'original process' $(X_n, \Gamma_n)_{n\geq 1}$ from Section 3 and the random variables $(Y_n, U_n)_{n\geq 2}$ following (12) and (13), so that the two processes coincide in the set $\bigcap_{n=1}^\infty \{\Gamma_n \in K_n\}$.

Before stating the ergodicity result from [21] for this truncated chain, four technical assumptions are listed, which must hold for some constants $c \geq 1$ and $\beta \geq 0$ and $\iota \in (0, \frac{1}{2})$.

(A1) For all measurable $A \subset \mathbb{X}$, it holds that $\mathbb{P}(\tilde{X}_{n+1} \in A \mid \mathcal{F}_n) = P_{\tilde{\Gamma}_n}(\tilde{X}_n, A)$ almost surely, and for each $\gamma \in \mathbb{G}$, the transition probability P_γ has π as the unique invariant distribution.

(A2) For each $n \geq 1$, the following uniform drift and minorisation conditions hold for all $\gamma \in K_n$, for all $x \in \mathbb{X}$ and all measurable $A \subset \mathbb{X}$

$$P_\gamma V(x) \leq \lambda_n V(x) + b_n \mathbb{1}_{C_n}(x)$$
$$P_\gamma(x, A) \geq \delta_n \mathbb{1}_{C_n}(x)\nu_\gamma(A)$$

where $C_n \subset \mathbb{X}$ is a subset (a minorisation set), $V : \mathbb{X} \to [1, \infty)$ is a drift function such that $\sup_{x\in C_n} V(x) \leq b_n$ and ν_γ 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, $\delta_n \in (0, 1]$ is decreasing with respect to n and they are polynomially bounded so that

$$\max\{(1-\lambda_n)^{-1}, \delta_n^{-1}, b_n\} \leq cn^\beta.$$

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

$$\|P_\gamma f - P_{\gamma'} f\|_{V^r} \leq c' n^\beta \|f\|_{V^r} |\gamma - \gamma'|$$

with the norm on the space $\bar{\mathbb{G}}$ defined as $|\gamma| = |(s, \mu, \Sigma)| = |s| + \|\mu\| + \|\Sigma\|$.

(A4) The inequality $|\hat{H}(\gamma, x, y)| \leq c n^\beta V^\iota(x)$ holds for all $\gamma \in K_n$ and all $x, y \in \mathbb{X}$.

Theorem 22. *Assume* (A1)–(A4) *hold and let* f *be a function with* $\|f\|_{V^\tau} < \infty$ *for some* $\tau \in (0, 1 - \iota)$. *Assume* $\beta < \kappa_*^{-1} \min\{1/2, 1 - \iota - \tau\}$ *and* $\sum_{k=1}^\infty k^{\kappa_* \beta - 1} \eta_k < \infty$ *where* $\kappa_* \geq 1$ *is an independent constant. Then,*

$$\frac{1}{n} \sum_{k=1}^n f(\tilde{X}_k) \xrightarrow{n \to \infty} \int_{\mathbb{X}} f(x) \pi(x) \mathrm{d}x \quad \textit{almost surely.} \tag{26}$$

Proof. This theorem is a straightforward modification of Theorem 2 in [21]. In particular, the assumption (A4) here is only slightly more general than assumption (A4) in [21] and the changes required for the proof are obvious. $\square$

Now we are ready to give a proof to the first main result considering the case of compactly supported π.

Proof of Theorem 7. Corollary 19 ensures that for any $\epsilon > 0$, there are $-\infty < a_1^{(\epsilon)} < a_2^{(\epsilon)} < \infty$ such that $\mathbb{P}(B^{(\epsilon)}) \geq 1 - \epsilon$, where

$$B^{(\epsilon)} := \{a_1^{(\epsilon)} \leq S_n \leq a_2^{(\epsilon)} \text{ for all } n \geq 1\}.$$

Set $K_n^{(\epsilon)} := K^{(\epsilon)} := [a_1^{(\epsilon)}, a_2^{(\epsilon)}] \times \mathbb{S}_\zeta$ for all $n \geq 1$, and construct the truncated process $(\tilde{X}_n^{(\epsilon)}, \tilde{\Gamma}_n^{(\epsilon)})$ using these restriction sets in (25). Define $\theta_1^{(\epsilon)} := \phi(a_1^{(\epsilon)}) > 0$ and $\theta_2^{(\epsilon)} := \phi(a_2^{(\epsilon)}) < \infty$.

Let us next verify the above assumptions (A1)–(A4) with some $c \geq 1$, $\beta = 0$ and $V \equiv 1$. The assumption (A1) holds by construction of the process and the Metropolis kernel. For (A2), take $C_n := \mathbb{X}$ for all $n \geq 1$, and notice that $P_\gamma V(x) = 1$ for all $x \in \mathbb{X}$ and $\gamma \in \mathbb{G}$. By Assumption 2 one can estimate for all $\gamma \in K^{(\epsilon)}$ and all $x \in \mathbb{X}$,

$$\begin{aligned}
P_\gamma(x, A) &\geq \int_A \alpha(x, y) q_\gamma(x - y) \mathrm{d}y \\
&\geq \left(\inf_{x, y \in \mathbb{X},\, \gamma \in K^{(\epsilon)}} q_\gamma(x - y) \right) \int_A \frac{\pi(y)}{\sup_{z \in \mathbb{X}} \pi(z)} \mathrm{d}y \\
&\geq \theta_2^{-d} \zeta^{-1/2} \left(\inf_{|z| \leq \operatorname{diam}(\mathbb{X})} \hat{q}(\|\theta_1^{-1} \zeta^{1/2} z\|) \right) c_1 \nu_\gamma(A) \geq \delta \nu_\gamma(A)
\end{aligned}$$

with a $\delta > 0$, where $\nu_\gamma(A) := \nu(A) := c_1^{-1} \int_A \frac{\pi(y)}{\sup_{z \in \mathbb{X}} \pi(z)} \mathrm{d}y$ and $c_1 > 0$ chosen so that $\nu(\mathbb{X}) = 1$. Assumption 1 ensures that the derivative of ϕ is bounded on $[a_1^{(\epsilon)}, a_2^{(\epsilon)}]$ and therefore we have

$$\|\phi(s)\Sigma^{1/2} - \phi(s')\Sigma'^{1/2}\| \leq \|\Sigma\| \cdot |\phi(s) - \phi(s')| + |\phi(s)| \cdot \|\Sigma - \Sigma'\| \leq c_2 |\gamma - \gamma'|$$

with some finite $c_2 = c_2(\epsilon)$ and Proposition 26 in Appendix B implies (A3). Finally, it holds that $|H(\gamma, x, y)| \le c$ for all $\gamma \in K_n$ and $x, y \in \mathbb{X}$, implying (A4).

All (A1)–(A4) hold and $\sum_{k=1}^{\infty} k^{-1}\eta_k \le (\sum_{k=1}^{\infty} k^{-2})^{1/2}(\sum_{k=1}^{\infty} \eta_k^2)^{1/2} < \infty$ by Assumption 3, so Theorem 22 yields a strong law of large numbers for the truncated process $\tilde{X}_n^{(\epsilon)}$ in case of a bounded function f. Since $(\tilde{X}_n^{(\epsilon)})_{n\ge1}$ coincides with the original process $(X_n)_{n\ge1}$ in $B^{(\epsilon)}$, the ergodic averages corresponding to $X_n(\omega)$ converge to $\int f(x)\pi(x)\mathrm{d}x$ with almost every $\omega \in B^{(\epsilon)}$. Since $\epsilon > 0$ was arbitrary, the strong law of large numbers (8) holds almost surely. $\square$

Remark 23. Theorem 22 (Theorem 2 of [21]) is a modification of Proposition 6 in [1]. Having Corollary 19 ensuring the boundedness of the trajectories of S_n, Theorem 7 could be obtained also using other techniques, in particular, the mixingale approach described in [6,11], or the coupling technique of [19] (resulting in a weak law of large numbers). These other techniques do not, however, apply directly to Theorem 9, since in this case the trajectories of S_n are not necessarily bounded from above, but only satisfy the polynomial bound of Proposition 20.

Proof of Theorem 9. Proposition 20 ensures that for any $\beta' > 0$ there are a.s. positive Θ_1 and a.s. finite Θ_2 such that

$$\Theta_1 \le \phi(S_n) \le \Theta_2 n^{\beta'}. \tag{27}$$

Now, similarly as in the proof of Theorem 7, for any $\epsilon > 0$, one can find $0 < \theta_1^{(\epsilon)} \le \theta_2^{(\epsilon)} < \infty$ such that

$$\mathbb{P}(\forall n \ge 1 : \theta_1^{(\epsilon)} \le \phi(S_n) \le \theta_2^{(\epsilon)} n^{\beta'}) \ge 1 - \epsilon \tag{28}$$

and construct $(\tilde{X}_n^{(\epsilon)}, \tilde{S}_n^{(\epsilon)})_{n\ge1}$ using the restriction sets $K_n^{(\epsilon)} := [a_1^{(\epsilon)}, a_2^{(n,\epsilon)}]$, where $\phi(a_1^{(\epsilon)}) = \theta_1^{(\epsilon)}$ and $\phi(a_2^{(\epsilon,n)}) = \theta_2^{(\epsilon)} n^{\beta'}$.

Let $\xi \in (p, 1)$ and let $V(x) := c_V \pi^{-\xi}(x)$ with $c_V := \sup_x \pi^{\xi}(x)$. Assumption (A1) holds by construction and (A4) holds for any given $\iota \in (0, 1-\xi)$ as verified in the proof of Theorem 10 in [21], observing that $|H(x, y)| \le 1$. Proposition 25 in Appendix B with the fact $\det(\theta\Sigma) = \theta^d \det(\Sigma)$ yields (A2) with $\beta = d\beta'$. Assumption 1 ensures that $\phi'(s) \le c_1\phi^{\kappa}(s)$ for all $s \in \mathbb{R}$, from which $|\phi(s) - \phi(s')| \le c_1(\theta_2^{(\epsilon)} n^{\beta'})^{\kappa}|s - s'| \le c_2 n^{\kappa\beta'}|s - s'|$ for all $s, s' \in [a_1^{(\epsilon)}, a_2^{(n,\epsilon)}]$. Now, Proposition 26 in Appendix B shows (A3) with $\beta = c_3\beta'$ as in the proof of Theorem 7. To conclude, the assumptions (A1)–(A4) hold with constants (c, β), where $\beta = \beta(\epsilon, \beta') > 0$ can be selected to be arbitrarily small and $c = c(\epsilon, \beta) < \infty$.

In particular, one can let $\beta < 1/2\kappa_*^{-1}$, so that $\sum_{k=1}^{\infty} k^{\kappa_*\beta-1}\eta_k < \infty$ as in the proof of Theorem 7. Take now $\tau = p/\xi \in (0, 1)$ so that $|f(x)|/V^{\tau}(x) = c_V^{\tau}|f(x)|\pi^{p}(x)$, implying that $\|f\|_{V^{\tau}} < \infty$. Theorem 22 guarantees that the strong law of large numbers holds in the set (28), and a.s. by letting $\epsilon \to 0$. $\square$

Acknowledgments

The author thanks Professor Eero Saksman and the anonymous reviewers for comments significantly improving the presentation of the paper. The author was 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.

Appendix A. Proofs of geometric lemmas

Proof of Lemma 15. Let $\Sigma \in \mathbb{S}_\zeta$ with $\zeta \in [1, \infty)$, that is, the set of eigenvalues satisfy $\lambda(\Sigma) \subset [\zeta^{-1}, \zeta]$. Then $\zeta^{-d} \le \det(\Sigma) \le \zeta^d$ and the claim (18) follows by

$$\sup_{\Sigma\in\mathbb{S}_\zeta,\, z\in\mathbb{R}^d} q_{(s,\Sigma)}(z) \le [\phi(s)]^{-d}\zeta^{d/2} \sup_{z\in\mathbb{R}^d} q(z).$$

Observe then that for any constant $M > 0$ one has

$$\int_{\overline{B}(0,\phi(s)M)} [\phi(s)]^{-d} \det(\Sigma)^{-1/2} q([\phi(s)]^{-1}\Sigma^{-1/2}z)\mathrm{d}z \ge \int_{\overline{B}(0,\zeta^{-1/2}M)} q(u)\mathrm{d}u$$

since $u \in \overline{B}(0, \xi^{-1/2}M)$ implies that $[\phi(s)]\Sigma^{1/2}u \in \overline{B}(0, \phi(s)M)$. Clearly M can be chosen sufficiently large so that (19) holds.

Let then $P \subset \mathbb{R}^d$ be a plane, and let $z \in \mathbb{R}^d$ such that $d(z, P) \le \phi(s)M^{-1}$. Denote by z^* the orthogonal projection of z to P, whence $\|z^* - z\| \le \phi(s)M^{-1}$. Denote then $u = [\phi(s)]^{-1}\Sigma^{-1/2}z$ and $u^* = [\phi(s)]^{-1}\Sigma^{-1/2}z^*$. We obtain that

$$\|u - u^*\| \le [\phi(s)]^{-1}\zeta^{1/2}\|z - z^*\| \le \zeta^{1/2}M^{-1}.$$

Having this estimate, we can estimate

$$\int_{\{d(z,P)\le\phi(s)M^{-1}\}} q_{(s,\Sigma)}(z)\mathrm{d}z \le \int_{\{d(u,\tilde{P})\le\zeta^{1/2}M^{-1}\}} q(u)\mathrm{d}u$$

where $\tilde{P} = [\phi(s)]^{-1}\Sigma^{-1/2}P$ is a plane. To conclude, we may choose M sufficiently large so that (20) and (19) hold. □

Proof of Lemma 18. Fix an $\epsilon' > 0$. By the uniform smoothness of $\{\partial A_i\}_{i\in I}$, one can find $\delta > 0$ so that $\|n_i(y) - n_i(z)\| \le \epsilon'$ for all $i \in I$ and $y, z \in \partial A_i$ with $\|y - z\| \le 2\delta$.

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

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

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

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

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

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

Let us define the half-space $T := \{y - 2\epsilon' r n_i(y) + z : z \cdot n_i(y) < 0\}$. It holds that $\overline{B}(y, 2r) \cap T \subset \overline{B}(y, 2r) \cap C_-$ since taking $y + w \in \overline{B}(y, 2r) \cap T$ one has $n_i(y) \cdot w <$

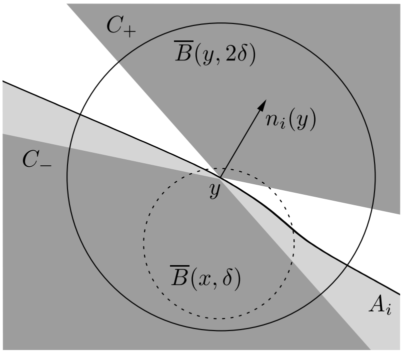

Fig. A.1. Illustration of the half-space approximation. The set A_i is shown in light grey, and the cones C_- and C_+ in dark grey.

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

Appendix B. Simultaneous properties for Metropolis kernels

We shall consider here the following general assumption on the proposal densities.

Assumption 24. Let $\mathcal{C}_d \subset \mathbb{R}^{d\times d}$ stand for the symmetric and positive definite matrices. Suppose $\mathcal{P} \subset \mathcal{C}_d$ and $\{q_R\}_{R\in\mathcal{P}}$ is a family of probability densities defined through

$$q_R(z) := |\det(R)|^{-1} \hat{q}(\|R^{-1} z\|), \tag{B.1}$$

where $\hat{q} : [0, \infty) \to (0, \infty)$ is a bounded, decreasing, and differentiable function, satisfying the conditions in Assumption 2. Moreover, suppose that there is a constant $\kappa > 0$ such that all the eigenvalues of each $R \in \mathcal{P}$ are bounded from below by κ.

Proposition 25. *Suppose π satisfies Assumption 8 and the family $\{q_R\}_{R\in\mathcal{P}}$ satisfies Assumption 24 with some $\kappa > 0$ and $\beta \in (0, 1)$. Let P_R be the Metropolis transition probability defined in (15) and using the proposal density q_R. 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 all $R \in \mathcal{P}$, $x \in \mathbb{R}^d$ and measurable $A \subset \mathbb{R}^d$,*

$$P_R V(x) \leq \lambda_R V(x) + b \mathbb{1}_C(x) \tag{B.2}$$

$$P_R(x, A) \geq \delta_R \mathbb{1}_C(x) \nu(A) \tag{B.3}$$

where $V(x) := c_V \pi^{-\beta}(x) \geq 1$ with $c_V := \sup_x \pi^\beta(x)$ and the constants $\lambda_R, \delta_R \in (0, 1)$ satisfy the bound

$$\max\{(1 - \lambda_R)^{-1}, \delta_R^{-1}\} \leq c |\det(R)|^{-1}$$

for some constant $c \geq 1$.

Proof. Proposition 25 is a generalisation of [21, Proposition 15] considering Gaussian densities q_R and the case $\beta = 1/2$. We shall describe the changes in the proof of [21, Proposition 15] required for the class of proposal distributions in Assumption 24.

First, observe that with $V(x) = c_V \pi^{-\beta}(x)$ one has

$$1 - \frac{P_R V(x)}{V(x)} = \int_{A_x} \left[1 - \left(\frac{\pi(x)}{\pi(y)}\right)^{\beta}\right] q_R(y-x)\mathrm{d}y \\ - \int_{R_x} \left(\frac{\pi(y)}{\pi(x)}\right)^{1-\beta} \left[1 - \left(\frac{\pi(y)}{\pi(x)}\right)^{\beta}\right] q_R(y-x)\mathrm{d}y.$$

The 1/4 in the estimate (37) of [21] is replaced with $c_* = \sup_{u\in[0,1]} u^{1-\beta}(1-u^\beta) \in (0,1)$. One can easily make $1 - (\pi(x)/\pi(y))^\beta > c^*$ for all $y \in \tilde{A}_x$, where c^* is any chosen value in $(c_*, 1)$.

For a non-negative function f, one can write by Fubini's theorem

$$\int_{\mathbb{R}^d} f(z+x) q_R(z)\mathrm{d}z = |\det(R)|^{-1} \int_0^{\hat{q}(0)} \int_{\{\hat{q}(\|R^{-1}z\|)\geq t\}} f(z+x)\mathrm{d}z\mathrm{d}t \\ = -|\det(R)|^{-1} \int_0^\infty \int_{E_u} f(y)\mathrm{d}y \hat{q}'(u)\mathrm{d}u$$

where the substitution $t = \hat{q}(u)$ was used, and $E_u := \{x + z : \|R^{-1}z\| \leq u\}$. One has $\|R^{-1}z\| \leq \kappa^{-1}\|z\|$, and thus $E_u \supset \overline{B}(x, u\kappa)$.

The conditions in Assumption 2 for the derivative $\hat{q}'$ correspond to the estimate obtained in [21, Lemma 14] for a Gaussian family, that is, $\hat{q} = e^{-x^2/2}$ and the case $\xi = 1/2$. In the present case, the choice $\xi = c_*/c^*$ is used. These facts are enough to complete the proof of [21, Proposition 15] to yield the claim. □

Proposition 26. *Suppose the family $\{q_R\}_{R\in\mathcal{P}}$ satisfies Assumption 24 with some $\kappa > 0$. Suppose, in addition, that either*

(i) $V \equiv 1$ *or*
(ii) π *satisfies Assumption 8 and* $\beta \in (0,1)$, $V(x) := c_V \pi^{-\beta}(x) \geq 1$ *with* $c_V := \sup_x \pi^\beta(x)$.

Then, there is a constant $c > 0$ such that for the Metropolis transition probability P_R given in (15), it holds that

$$\|P_R f - P_{R'} f\|_{V^r} \leq c \max\{\|R\|, \|R'\|\}^{d+1} \|f\|_{V^r} \|R - R'\| \tag{B.4}$$

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

Proof. Consider first (i). From the definition of the Metropolis kernel (15), one obtains

$$\sup_x |P_R f(x) - P_{R'} f(x)| \leq 2 \sup_x |f(x)| \int_{\mathbb{X}} |q_R(x) - q_{R'}(x)|\mathrm{d}x.$$

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

$$\|P_R f - P_{R'} f\|_{V^r} \leq 2\|f\|_{V^r} \int_{\mathbb{R}^d} |q_R(x) - q_{R'}(x)|\mathrm{d}x$$

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

As in [1,11], one can write

$$\int_{\mathbb{X}} |q_R(x) - q_{R'}(x)|\mathrm{d}x = \int_{\mathbb{X}} \left| \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} q_{R_t}(x)\mathrm{d}t \right| \mathrm{d}x$$

where $R_t := R' + t(R - R')$. Let us compute

$$\frac{\mathrm{d}}{\mathrm{d}t} q_{R_t}(x) = -\mathrm{tr}\big(R_t^{-1}(R - R')\big) q_{R_t}(x) + |\det(R_t)|^{-1} \hat{q}'(\|R_t^{-1}x\|) \frac{\mathrm{d}}{\mathrm{d}t}\|R_t^{-1}x\|$$

and

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

Since $R - R'$ and R_t^{-1} are symmetric and R_t^{-1} positive definite, it holds that $|\mathrm{tr}(R_t^{-1}(R - R'))| \le \mathrm{tr}(R_t^{-1}) \max_{1\le i\le d} |\lambda_i| \le \mathrm{tr}(R_t^{-1})\|R - R'\|$ where λ_i are the eigenvalues of $R - R'$ (see, e.g, [24]). Since the Frobenius norm is sub-multiplicative,

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

by polar integration. Denote $\lambda = \lambda(u, t) := \|R_t^{-1}u\|$, and observe that since $\hat{q}$ is decreasing, integration by parts yields

$$\begin{aligned}
\int_0^M r^d |\hat{q}'(\lambda r)|\mathrm{d}r &= \frac{d}{\lambda}\int_0^M r^{d-1}\hat{q}(\lambda r)\mathrm{d}r - M^d \frac{\hat{q}(\lambda M)}{\lambda} \\
&\le \frac{d}{\lambda^{d+1}} \int_0^\infty u^{d-1}\hat{q}(u)\mathrm{d}u = \frac{d c_{\hat{q}}}{\lambda^{d+1}}
\end{aligned}$$

for all $M > 0$. Since λ^{-1} is smaller, for any $\|u\| = 1$ and $t \in [0, 1]$, than the maximum eigenvalue of R and R', which is smaller than $\max\{\|R\|, \|R'\|\}$, we obtain

$$\int_{\mathbb{R}^d} |q_R(x) - q_{R'}(x)|\mathrm{d}x \le c_1 \max\{\|R\|, \|R'\|\}^{d+1}\|R - R'\|$$

concluding the proof with $c = 2c_1$. □

Proposition 27. *Suppose the proposal density q is given as $q(z) = c\tilde{q}(\|z\|)$ where $c > 0$ is a constant and*

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

That is, q is a (multivariate) Gaussian or Student distribution, respectively. Then, q satisfies Assumption 2.

Proof. It is sufficient to verify that the derivative of $\tilde{q}$ satisfies the conditions in Assumption 2. Fix $\xi \in (0, 1)$ and assume $\epsilon > 0$. Consider first (i), in which case

$$\begin{aligned}\xi\tilde{q}'(x) - \tilde{q}'(x+\epsilon) &= (x+\epsilon)e^{-(x+\epsilon)^2/2} - \xi x e^{-x^2/2} \\ &\geq x e^{-x^2/2}[e^{-\epsilon x - \epsilon^2/2} - \xi] > 0\end{aligned}$$

if and only if $x < x_\epsilon := -\frac{\epsilon}{2} - \frac{\log\xi}{\epsilon}$. Let $\epsilon_* \in (0, 1)$ be small enough so that $x_\epsilon > 0$ for all $\epsilon \in (0, \epsilon_*]$, from which one obtains $c_1 > 0$ and $0 \leq a < b < \infty$ such that $\xi\tilde{q}'(x) - \tilde{q}'(x+\epsilon) \geq c_1$ for all $x \in [a, b]$ and all $\epsilon \in [0, \epsilon_*]$. Moreover, for all $\epsilon \in (0, \epsilon_*)$

$$\begin{aligned}\int_0^\infty \min\{0, \xi\tilde{q}'(x) - \tilde{q}'(x+\epsilon)\}\mathrm{d}x &\geq \int_{x_\epsilon}^\infty x e^{-x^2/2}[e^{-\epsilon x - \epsilon^2/2} - \xi]\mathrm{d}x \geq -\xi e^{-x_\epsilon^2/2} \\ &= -\xi e^{-\epsilon^2/8 - \log(\xi)/2} e^{-(\log\xi)^2\epsilon^{-2}/2} \geq -c_2 e^{-c_3\epsilon^{-1}}\end{aligned}$$

with $c_2 = \xi e^{-\log(\xi)/2}$ and $c_3 = (\log\xi)^2/2$.

Assume then (ii). By the mean value theorem, denoting $c := d + 2p$ and $\alpha := d/2 + p + 1$, one can write for some $\epsilon' \in [0, \epsilon]$

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

for all $x > 0$, whenever $\epsilon > 0$ is sufficiently small. The claim follows easily. □

References

[1] C. Andrieu, É Moulines, On the ergodicity properties of some adaptive MCMC algorithms, Ann. Appl. Probab. 16 (2006) 1462–1505.

[2] C. Andrieu, É Moulines, P. Priouret, Stability of stochastic approximation under verifiable conditions, SIAM J. Control Optim. 44 (2005) 283–312.

[3] C. Andrieu, C.P. Robert, Controlled MCMC for Optimal Sampling, Technical Report Ceremade 0125, Université Paris Dauphine, 2001.

[4] C. Andrieu, J. Thoms, A tutorial on adaptive MCMC, Statist. Comput. 18 (2008) 343–373.

[5] Y. Atchadé, G. Fort, Limit theorems for some adaptive MCMC algorithms with subgeometric kernels, Bernoulli 16 (2010) 116–154.

[6] Y.F. Atchadé, J.S. Rosenthal, On adaptive Markov chain Monte Carlo algorithms, Bernoulli 11 (2005) 815–828.

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

[8] G. Fort, E. Moulines, P. Priouret, Convergence of adaptive MCMC algorithms: ergodicity and law of large numbers, Technical Report, 2010.

[9] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, Markov Chain Monte Carlo in Practice, Chapman & Hall/CRC, Boca Raton, Florida, 1998.

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

[11] H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis algorithm, Bernoulli 7 (2001) 223–242.

[12] P. Hall, C.C. Heyde, Martingale Limit Theory and its Application, Academic Press, New York, 1980.

[13] S.F. Jarner, E. Hansen, Geometric ergodicity of Metropolis algorithms, Stochastic Process. Appl. 85 (2000) 341–361.

[14] E. Nummelin, MC's for MCMC'ists, Internat. Statist. Rev. 70 (2002) 215–240.

[15] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer-Verlag, New York, 1999.
[16] G.O. Roberts, A. Gelman, W.R. Gilks, Weak convergence and optimal scaling of random walk Metropolis algorithms, Ann. Appl. Probab. 7 (1997) 110–120.
[17] G.O. Roberts, J.S. Rosenthal, Optimal scaling for various Metropolis–Hastings algorithms, Statist. Sci. 16 (2001) 351–367.
[18] G.O. Roberts, J.S. Rosenthal, General state space Markov chains and MCMC algorithms, Probab. Surv. 1 (2004) 20–71.
[19] G.O. Roberts, J.S. Rosenthal, Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms, J. Appl. Probab. 44 (2007) 458–475.
[20] G.O. Roberts, J.S. Rosenthal, Examples of adaptive MCMC, J. Comput. Graph. Statist. 18 (2009) 349–367.
[21] E. Saksman, M. Vihola, On the ergodicity of the adaptive Metropolis algorithm on unbounded domains, Ann. Appl. Probab. 20 (2010) 2178–2203.
[22] C. Sherlock, G. Roberts, Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets, Bernoulli 15 (2009) 774–798.
[23] M. Vihola, Can the adaptive Metropolis algorithm collapse without the covariance lower bound? Electron. J. Probab. 16 (2011) 45–75.
[24] S.D. Wang, T.S. Kuo, C.F. Hsu, Trace bounds on the solution of the algebraic matrix Riccati and Lyapunov equation, IEEE Trans. Automat. Control 31 (1986) 654–656.