DOI: 10.1214/EJP.v16-840

Electronic Journal of Probability

Vol. 16 (2011), Paper no. 2, pages 45–75.

Journal URL
http://www.math.washington.edu/~ejpecp/

Can the adaptive Metropolis algorithm collapse without the covariance lower bound?*

Matti Vihola
Department of Mathematics and Statistics
University of Jyväskylä
P.O.Box 35
FI-40014 University of Jyväskyla
Finland
matti.vihola@iki.fi
http://iki.fi/mvihola/

Abstract

The Adaptive Metropolis (AM) algorithm is based on the symmetric random-walk Metropolis algorithm. The proposal distribution has the following time-dependent covariance matrix at step $n+1$

$$S_n = \mathrm{Cov}(X_1,\ldots,X_n) + \epsilon I,$$

that is, the sample covariance matrix of the history of the chain plus a (small) constant $\epsilon > 0$ multiple of the identity matrix I. The lower bound on the eigenvalues of S_n induced by the factor ϵI is theoretically convenient, but practically cumbersome, as a good value for the parameter ϵ may not always be easy to choose. This article considers variants of the AM algorithm that do not explicitly bound the eigenvalues of S_n away from zero. The behaviour of S_n is studied in detail, indicating that the eigenvalues of S_n do not tend to collapse to zero in general. In dimension one, it is shown that S_n is bounded away from zero if the logarithmic target density is uniformly continuous. For a modification of the AM algorithm including an additional fixed

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

component in the proposal distribution, the eigenvalues of S_n are shown to stay away from zero with a practically non-restrictive condition. This result implies a strong law of large numbers for super-exponentially decaying target distributions with regular contours.

Key words: Adaptive Markov chain Monte Carlo, Metropolis algorithm, stability, stochastic approximation.

AMS 2000 Subject Classification: Primary 65C40.

Submitted to EJP on January 14, 2010, final version accepted November 27, 2011.

1 Introduction

Adaptive Markov chain Monte Carlo (MCMC) methods have attracted increasing interest in the last few years, after the original work of Haario, Saksman, and Tamminen Haario et al. (2001) and the subsequent advances in the field Andrieu and Moulines (2006); Andrieu and Robert (2001); Atchadé and Rosenthal (2005); Roberts and Rosenthal (2007); see also the recent review Andrieu and Thoms (2008). Several adaptive MCMC algorithms have been proposed up to date, but the seminal Adaptive Metropolis (AM) algorithm Haario et al. (2001) is still one of the most applied methods, perhaps due to its simplicity and generality.

The AM algorithm is a symmetric random-walk Metropolis algorithm, with an adaptive proposal distribution. The algorithm starts[1] at some point $X_1 \equiv x_1 \in \mathbb{R}^d$ with an initial positive definite covariance matrix $S_1 \equiv s_1 \in \mathbb{R}^{d\times d}$ and follows the recursion

(S1) Let $Y_{n+1} = X_n + \theta S_n^{1/2} W_{n+1}$, where W_{n+1} is an independent standard Gaussian random vector and $\theta > 0$ is a constant.

(S2) Accept Y_{n+1} with probability $\min\left\{1, \frac{\pi(Y_{n+1})}{\pi(X_n)}\right\}$ and let $X_{n+1} = Y_{n+1}$; otherwise reject Y_{n+1} and let $X_{n+1} = X_n$.

(S3) Set $S_{n+1} = \Gamma(X_1, \ldots, X_{n+1})$.

In the original work Haario et al. (2001) the covariance parameter is computed by

$$\Gamma(X_1,\ldots,X_{n+1}) = \frac{1}{n}\sum_{k=1}^{n+1}(X_k - \overline{X}_{n+1})(X_k - \overline{X}_{n+1})^T + \epsilon I, \tag{1}$$

where $\overline{X}_n := n^{-1}\sum_{k=1}^{n} X_k$ stands for the mean. That is, S_{n+1} is a covariance estimate of the history of the 'Metropolis chain' $X_1, \ldots, X_{n+1}$ plus a small $\epsilon > 0$ multiple of the identity matrix $I \in \mathbb{R}^{d\times d}$. The authors prove a strong law of large numbers (SLLN) for the algorithm, that is, $n^{-1}\sum_{k=1}^{n} f(X_k) \to \int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x$ almost surely as $n\to\infty$ for any bounded functional f when the target distribution π is bounded and compactly supported. Recently, SLLN was shown to hold also for π with unbounded support, having super-exponentially decaying tails with regular contours and f growing at most exponentially in the tails Saksman and Vihola (2010).

This article considers the original AM algorithm (S1)–(S3), without the lower bound induced by the factor ϵI. The proposal covariance function Γ, defined precisely in Section 2, is a consistent covariance estimator first proposed in Andrieu and Robert (2001). A special case of this estimator behaves asymptotically like the sample covariance in (1). Previous results indicate that if this algorithm is modified by truncating the eigenvalues of S_n within explicit lower and upper bounds, the algorithm can be verified in a fairly general setting Atchadé and Fort (2010); Roberts and Rosenthal (2007). It is also possible to determine an increasing sequence of truncation sets for S_n, and modify the algorithm to include a re-projection scheme in order to verify the validity of the algorithm Andrieu and Moulines (2006).

While technically convenient, such pre-defined bounds on the adapted covariance matrix S_n can be inconvenient in practice. Ill-defined values can affect the efficiency of the adaptive scheme dramatically, rendering the algorithm useless in the worst case. In particular, if the factor $\epsilon > 0$ in the AM

[1] The initial 'burn-in' phase included in the original algorithm is not considered here.

algorithm is selected too large, the smallest eigenvalue of the true covariance matrix of π may be well smaller than $\epsilon > 0$, and the chain X_n is likely to mix poorly. Even though the re-projection scheme of Andrieu and Moulines (2006) avoids such behaviour by increasing truncation sets, which eventually contain the desirable values of the adaptation parameter, the practical efficiency of the algorithm is still strongly affected by the choice of these sets Andrieu and Thoms (2008).

Without a lower bound on the eigenvalues of S_n (or a re-projection scheme), there is a potential danger of the covariance parameter S_n collapsing to singularity. In such a case, the increments $X_n - X_{n-1}$ would be smaller and smaller, and the X_n chain could eventually get 'stuck'. The empirical evidence suggests that this does not tend to happen in practice. The present results validate the empirical findings by excluding such a behaviour in different settings.

After defining precisely the algorithms in Section 2, the above mentioned unconstrained AM algorithm is analysed in Section 3. First, the AM algorithm run on an improper uniform target $\pi \equiv c > 0$ is studied. In such a case, the asymptotic expected growth rate of S_n is characterised quite precisely, being $e^{2\theta\sqrt{n}}$ for the original AM algorithm Haario et al. (2001). The behaviour of the AM algorithm in the uniform target setting is believed to be similar as in a situation where S_n is small and the target π is smooth whence locally constant. The results support the strategy of choosing a 'small' initial covariance s_1 in practice, and letting the adaptation take care of expanding it to the proper size.

In Section 3, it is also shown that in a one-dimensional setting and with a uniformly continuous $\log \pi$, the variance parameter S_n is bounded away from zero. This fact is shown to imply, with the results in Saksman and Vihola (2010), a SLLN in the particular case of a Laplace target distribution. While this result has little practical value in its own right, it is the first case where the unconstrained AM algorithm is shown to preserve the correct ergodic properties. It shows that the algorithm possesses self-stabilising properties and further strengthens the belief that the algorithm would be stable and ergodic under a more general setting.

Section 4 considers a slightly different variant of the AM algorithm, due to Roberts and Rosenthal Roberts and Rosenthal (2009), replacing (S1) with

(S1') With probability β, let $Y_{n+1} = X_n + V_{n+1}$ where V_{n+1} is an independent sample of q_{fix}; otherwise, let $Y_{n+1} = X_n + \theta S_n^{1/2} W_{n+1}$ as in (S1).

While omitting the parameter $\epsilon > 0$, the proposal strategy (S1') includes two additional parameters: the mixing probability $\beta \in (0,1)$ and the fixed symmetric proposal distribution q_{fix}. It has the advantage that the 'worst case scenario' having ill-defined q_{fix} only 'wastes' the fixed proportion β of samples, while S_n can take any positive definite value on adaptation. This approach is analysed also in the recent preprint Bai et al. (2008), relying on a technical assumption that ultimately implies that X_n is bounded in probability. In particular, the authors show that if q_{fix} is a uniform density on a ball having a large enough radius, then the algorithm is ergodic. Section 4 uses a perhaps more transparent argument to show that the proposal strategy (S1') with a mild additional condition implies a sequence S_n with eigenvalues bounded away from zero. This fact implies a SLLN using the technique of Saksman and Vihola (2010), as shown in the end of Section 4.

2 The general algorithm

Let us define a Markov chain $(X_n, M_n, S_n)_{n\geq 1}$ evolving in space $\mathbb{R}^d \times \mathbb{R}^d \times \mathscr{C}^d$ with the state space $\mathbb{R}^d$ and $\mathscr{C}^d \subset \mathbb{R}^{d\times d}$ standing for the positive definite matrices. The chain starts at an initial position $X_1 \equiv x_1 \in \mathbb{R}^d$, with an initial mean[2] $M_1 \equiv m_1 \in \mathbb{R}^d$ and an initial covariance matrix $S_1 \equiv s_1 \in \mathscr{C}^d$. For $n \geq 1$, the chain is defined through the recursion

$$X_{n+1} \sim P_{q_{S_n}}(X_n, \cdot) \tag{2}$$
$$M_{n+1} := (1-\eta_{n+1})M_n + \eta_{n+1}X_{n+1} \tag{3}$$
$$S_{n+1} := (1-\eta_{n+1})S_n + \eta_{n+1}(X_{n+1}-M_n)(X_{n+1}-M_n)^T. \tag{4}$$

Denoting the natural filtration of the chain as $\mathscr{F}_n := \sigma(X_k, M_k, S_k : 1 \leq k \leq n)$, the notation in (2) reads that $\mathbb{P}\left(X_{n+1} \in A \,\middle|\, \mathscr{F}_n\right) = P_{q_{S_n}}(X_n, A)$ for any measurable $A \subset \mathbb{R}^d$. The Metropolis transition kernel P_q is defined for any symmetric probability density $q(x,y) = q(x-y)$ through

$$P_q(x,A) := \mathbb{1}_A(x)\left[1 - \int \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} q(y-x)\mathrm{d}y\right] + \int_A \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} q(y-x)\mathrm{d}y$$

where $\mathbb{1}_A$ stands for the characteristic function of the set A. The proposal densities $\{q_s\}_{s\in\mathscr{C}^d}$ are defined as a mixture

$$q_s(z) := (1-\beta)\tilde{q}_s(z) + \beta q_{\mathrm{fix}}(z) \tag{5}$$

where the mixing constant $\beta \in [0,1)$ determines the portion how often a fixed proposal density q_{fix} is used instead of the adaptive proposal $\tilde{q}_s(z) := \det(\theta s)^{-1/2}\tilde{q}(\theta^{-1/2}s^{-1/2}z)$ with $\tilde{q}$ being a 'template' probability density. Finally, the adaptation weights $(\eta_n)_{n\geq 2} \subset (0,1)$ appearing in (3) and (4) is assumed to decay to zero.

One can verify that for $\beta = 0$ this setting corresponds to the algorithm (S1)–(S3) of Section 1 with W_{n+1} having distribution $\tilde{q}$, and for $\beta \in (0,1)$, (S1') applies instead of (S1). Notice also that the original AM algorithm essentially fits this setting, with $\eta_n := n^{-1}$, $\beta := 0$ and if $\tilde{q}_s$ is defined slightly differently, being a Gaussian density with mean zero and covariance $s + \epsilon I$. Moreover, if one sets $\beta = 1$, the above setting reduces to a non-adaptive symmetric random walk Metropolis algorithm with the increment proposal distribution q_{fix}.

3 The unconstrained AM algorithm

3.1 Overview of the results

This section deals with the unconstrained AM algorithm, that is, the algorithm described in Section 2 with the mixing constant $\beta = 0$ in (5). Sections 3.2 and 3.3 consider the case of an improper uniform target distribution $\pi \equiv c$ for some constant $c > 0$. This implies that (almost) every proposed sample is accepted and the recursion (2) reduces to

$$X_{n+1} = X_n + \theta S_n^{1/2} W_{n+1} \tag{6}$$

[2] A customary choice is to set $m_1 = x_1$.

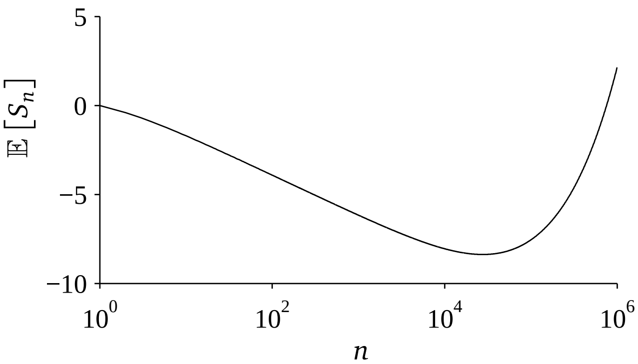

Figure 1: An example of the exact development of $\mathbb{E}\left[S_n\right]$, when $s_1 = 1$ and $\theta = 0.01$. The sequence $(\mathbb{E}\left[S_n\right])_{n\geq 1}$ decreases until n is over $27,000$ and exceeds the initial value only with n over $750,000$.

where $(W_n)_{n\geq 2}$ are independent realisations of the distribution $\tilde{q}$.

Throughout this subsection, let us assume that the template proposal distribution $\tilde{q}$ is spherically symmetric and the weight sequence is defined as $\eta_n := cn^{-\gamma}$ for some constants $c \in (0,1]$ and $\gamma \in (1/2, 1]$. The first result characterises the expected behaviour of S_n when $(X_n)_{n\geq 2}$ follows (6).

Theorem 1. *Suppose $(X_n)_{n\geq 2}$ follows the 'adaptive random walk' recursion* (6)*, with $\mathbb{E}W_n W_n^T = I$. Then, for all $\lambda > 1$ there is $n_0 \geq m$ such that for all $n \geq n_0$ and $k \geq 1$, the following bounds hold*

$$\frac{1}{\lambda}\left(\theta \sum_{j=n+1}^{n+k} \sqrt{\eta_j}\right) \leq \log\left(\frac{\mathbb{E}\left[S_{n+k}\right]}{\mathbb{E}\left[S_n\right]}\right) \leq \lambda\left(\theta \sum_{j=n+1}^{n+k} \sqrt{\eta_j}\right).$$

Proof. Theorem 1 is a special case of Theorem 12 in Section 3.2. □

Remark 2. Theorem 1 implies that with the choice $\eta_n := cn^{-\gamma}$ for some $c \in (0,1)$ and $\gamma \in (1/2, 1]$, the expectation grows with the speed

$$\mathbb{E}\left[S_n\right] \simeq \exp\left(\frac{\theta\sqrt{c}}{1-\frac{\gamma}{2}} n^{1-\frac{\gamma}{2}}\right).$$

Remark 3. In the original setting (Haario et al. 2001) the weights are defined as $\eta_n := n^{-1}$ and Theorem 1 implies that the asymptotic growth rate of $\mathbb{E}[S_n]$ is $e^{2\theta\sqrt{n}}$ when $(X_n)_{n\geq 2}$ follows (6). Suppose the value of S_n is very small compared to the scale of a smooth target distribution π. Then, it is expected that most of the proposal are accepted, X_n behaves almost as (6), and S_n is expected to grow approximately at the rate $e^{2\theta\sqrt{n}}$ until it reaches the correct magnitude. On the other hand, simple deterministic bound implies that S_n can decay slowly, only with the polynomial speed n^{-1}. Therefore, it may be safer to choose the initial s_1 small.

Remark 4. The selection of the scaling parameter $\theta > 0$ in the AM algorithm does not seem to affect the expected asymptotic behaviour S_n dramatically. However, the choice $0 < \theta \ll 1$ can result in an significant initial 'dip' of the adapted covariance values, as exemplified in Figure 1. Therefore, the values $\theta \ll 1$ are to be used with care. In this case, the significance of a successful burn-in is also emphasised.

It may seem that Theorem 1 would automatically also ensure that $S_n \to \infty$ also path-wise. This is not, however, the case. For example, consider the probability space $[0,1]$ with the Borel σ-algebra and the Lebesgue measure. Then $(M_n, \mathscr{F}_n)_{n\geq 1}$ defined as $M_n := 2^{2n}\mathbb{1}_{[0,2^{-n})}$ and $\mathscr{F}_n := \sigma(X_k : 1 \leq k \leq n)$ is, in fact, a submartingale. Moreover, $\mathbb{E}M_n = 2^n \to \infty$, but $M_n \to 0$ almost surely.

The AM process, however, does produce an unbounded sequence S_n.

Theorem 5. *Assume that $(X_n)_{n\geq 2}$ follows the 'adaptive random walk' recursion* (6). *Then, for any unit vector $u \in \mathbb{R}^d$, the process $u^T S_n u \to \infty$ almost surely.*

Proof. Theorem 5 is a special case of Theorem 18 in Section 3.3. □

In a one-dimensional setting, and when $\log \pi$ is uniformly continuous, the AM process can be approximated with the 'adaptive random walk' above, whenever S_n is small enough. This yields

Theorem 6. *Assume $d = 1$ and $\log \pi$ is uniformly continuous. Then, there is a constant $b > 0$ such that* $\liminf_{n\to\infty} S_n \geq b$.

Proof. Theorem 6 is a special case of Theorem 18 in Section 3.4. □

Finally, having Theorem 6, it is possible to establish

Theorem 7. *Assume $\tilde{q}$ is Gaussian, the one-dimensional target distribution is standard Laplace $\pi(x) := \frac{1}{2}e^{-|x|}$ and the functional $f : \mathbb{R} \to \mathbb{R}$ satisfies $\sup_x e^{-\gamma|x|}|f(x)| < \infty$ for some $\gamma \in (0, 1/2)$. Then, $n^{-1}\sum_{k=1}^n f(X_k) \to \int f(x)\pi(x)\mathrm{d}x$ almost surely as $n \to \infty$.*

Proof. Theorem 7 is a special case of Theorem 21 in Section 3.4. □

Remark 8. In the case $\eta_n := n^{-1}$, Theorem 7 implies that the parameters M_n and S_n of the adaptive chain converge to 0 and 2, that is, the true mean and variance of the target distribution π, respectively.

Remark 9. Theorem 6 (and Theorem 7) could probably be extended to cover also targets π with compact supports. Such an extension would, however, require specific handling of the boundary effects, which can lead to technicalities.

3.2 Uniform target: expected growth rate

Define the following matrix quantities

$$a_n := \mathbb{E}\left[(X_n - M_{n-1})(X_n - M_{n-1})^T\right] \quad (7)$$
$$b_n := \mathbb{E}\left[S_n\right] \quad (8)$$

for $n \geq 1$, with the convention that $a_1 \equiv 0 \in \mathbb{R}^{d\times d}$. One may write using (3) and (6)

$$X_{n+1} - M_n = X_n - M_n + \theta S_n^{1/2} W_{n+1} = (1-\eta_n)(X_n - M_{n-1}) + \theta S_n^{1/2} W_{n+1}.$$

If $\mathbb{E}W_n W_n^T = I$, one may easily compute

$$\begin{aligned}&\mathbb{E}\left[(X_{n+1} - M_n)(X_{n+1} - M_n)^T\right] \\ &\quad = (1-\eta_n)^2\,\mathbb{E}\left[(X_n - M_{n-1})(X_n - M_{n-1})^T\right] + \theta^2\mathbb{E}\left[S_n\right]\end{aligned}$$

since W_{n+1} is independent of $\mathscr{F}_n$ and zero-mean due to the symmetry of $\tilde{q}$. The values of $(a_n)_{n\geq 2}$ and $(b_n)_{n\geq 2}$ are therefore determined by the joint recursion

$$a_{n+1} = (1-\eta_n)^2 a_n + \theta^2 b_n \tag{9}$$
$$b_{n+1} = (1-\eta_{n+1}) b_n + \eta_{n+1} a_{n+1}. \tag{10}$$

Observe that for any constant unit vector $u \in \mathbb{R}^d$, the recursions (9) and (10) hold also for

$$a_{n+1}^{(u)} := \mathbb{E}\left[u^T (X_{n+1} - M_n)(X_{n+1} - M_n)^T u\right]$$
$$b_{n+1}^{(u)} := \mathbb{E}\left[u^T S_{n+1} u\right].$$

The rest of this section therefore dedicates to the analysis if the one-dimensional recursions (9) and (10), that is, $a_n, b_n \in \mathbb{R}_+$ for all $n \geq 1$. The first result shows that the tail of $(b_n)_{n\geq 1}$ is increasing.

Lemma 10. *Let $n_0 \geq 1$ and suppose $a_{n_0} \geq 0$, $b_{n_0} > 0$ and for $n \geq n_0$ the sequences a_n and b_n follow the recursions* (9) *and* (10)*, respectively. Then, there is a $m_0 \geq n_0$ such that $(b_n)_{n\geq m_0}$ is strictly increasing.*

Proof. If $\theta \geq 1$, we may estimate $a_{n+1} \geq (1-\eta_n)^2 a_n + b_n$ implying $b_{n+1} \geq b_n + \eta_{n+1}(1-\eta_n)^2 a_n$ for all $n \geq n_0$. Since $b_n > 0$ by construction, and therefore also $a_{n+1} \geq \theta^2 b_n > 0$, we have that $b_{n+1} > b_n$ for all $n \geq n_0 + 1$.

Suppose then $\theta < 1$. Solving a_{n+1} from (10) yields

$$a_{n+1} = \eta_{n+1}^{-1}\left(b_{n+1} - b_n\right) + b_n$$

Substituting this into (9), we obtain for $n \geq n_0 + 1$

$$\eta_{n+1}^{-1}\left(b_{n+1} - b_n\right) + b_n = (1-\eta_n)^2\left[\eta_n^{-1}\left(b_n - b_{n-1}\right) + b_{n-1}\right] + \theta^2 b_n$$

After some algebraic manipulation, this is equivalent to

$$b_{n+1} - b_n = \frac{\eta_{n+1}}{\eta_n}(1-\eta_n)^3(b_n - b_{n-1}) + \eta_{n+1}\left[(1-\eta_n)^2 - 1 + \theta^2\right] b_n. \tag{11}$$

Now, since $\eta_n \to 0$, we have that $(1-\eta_n)^2 - 1 + \theta^2 > 0$ whenever n is greater than some n_1. So, if we have for some $n' > n_1$ that $b_{n'} - b_{n'-1} \geq 0$, the sequence $(b_n)_{n\geq n'}$ is strictly increasing after n'.

Suppose conversely that $b_{n+1} - b_n < 0$ for all $n \geq n_1$. From (10), $b_{n+1} - b_n = \eta_{n+1}(a_{n+1} - b_n)$ and hence $b_n > a_{n+1}$ for $n \geq n_1$. Consequently, from (9), $a_{n+1} > (1-\eta_n)^2 a_n + \theta^2 a_{n+1}$, which is equivalent to

$$a_{n+1} > \frac{(1-\eta_n)^2}{1-\theta^2} a_n.$$

Since $\eta_n \to 0$, there is a $\mu > 1$ and n_2 such that $a_{n+1} \geq \mu a_n$ for all $n \geq n_2$. That is, $(a_n)_{n\geq n_2}$ grows at least geometrically, implying that after some time $a_{n+1} > b_n$, which is a contradiction. To conclude, there is an $m_0 \geq n_0$ such that $(b_n)_{n\geq m_0}$ is strictly increasing. □

Lemma 10 shows that the expectation $\mathbb{E}\left[u^T S_n u\right]$ is ultimately bounded from below, assuming only that $\eta_n \to 0$. By additional assumptions on the sequence η_n, the growth rate can be characterised in terms of the adaptation weight sequence.

Assumption 11. Suppose $(\eta_n)_{n\geq 1}\subset(0,1)$ and there is $m'\geq 2$ such that

(i) $(\eta_n)_{n\geq m'}$ is decreasing with $\eta_n\to 0$,

(ii) $(\eta_{n+1}^{-1/2}-\eta_n^{-1/2})_{n\geq m'}$ is decreasing and

(iii) $\sum_{n=2}^{\infty}\eta_n=\infty$.

The canonical example of a sequence satisfying Assumption 11 is the one assumed in Section 3.1, $\eta_n := cn^{-\gamma}$ for $c\in(0,1)$ and $\gamma\in(1/2,1]$.

Theorem 12. *Suppose $a_m\geq 0$ and $b_m>0$ for some $m\geq 1$, and for $n>m$ the a_n and b_n are given recursively by* (9) *and* (10)*, respectively. Suppose also that the sequence $(\eta_n)_{n\geq 2}$ satisfies Assumption 11 with some $m'\geq m$. Then, for all $\lambda>1$ there is $m_2\geq m'$ such that for all $n\geq m_2$ and $k\geq 1$, the following bounds hold*

$$\frac{1}{\lambda}\left(\theta\sum_{j=n+1}^{n+k}\sqrt{\eta_j}\right)\leq\log\left(\frac{b_{n+k}}{b_n}\right)\leq\lambda\left(\theta\sum_{j=n+1}^{n+k}\sqrt{\eta_j}\right).$$

Proof. Let m_0 be the index from Lemma 10 after which the sequence b_n is increasing. Let $m_1>\max\{m_0,m'\}$ and define the sequence $(z_n)_{n\geq m_1-1}$ by setting $z_{m_1-1}=b_{m_1-1}$ and $z_{m_1}=b_{m_1}$, and for $n\geq m_1$ through the recursion

$$z_{n+1}=z_n+\frac{\eta_{n+1}}{\eta_n}(1-\eta_n)^3(z_n-z_{n-1})+\eta_{n+1}\tilde{\theta}^2 z_n \tag{12}$$

where $\tilde{\theta}>0$ is a constant. Consider such a sequence $(z_n)_{n\geq m_1-1}$ and define another sequence $(g_n)_{n\geq m_1+1}$ through

$$\begin{aligned}g_{n+1}&:=\eta_{n+1}^{-1/2}\frac{z_{n+1}-z_n}{z_n}=\eta_{n+1}^{-1/2}\left[\frac{\eta_{n+1}}{\eta_n}(1-\eta_n)^3\frac{z_n-z_{n-1}}{z_{n-1}}\frac{z_{n-1}}{z_n}+\eta_{n+1}\tilde{\theta}^2\right]\\&=\eta_{n+1}^{1/2}\left(\frac{(1-\eta_n)^3}{\eta_n}\frac{g_n}{g_n+\eta_n^{-1/2}}+\tilde{\theta}^2\right).\end{aligned}$$

Lemma 33 in Appendix A shows that $g_n\to\tilde{\theta}$.

Let us consider next two sequences $(z_n^{(1)})_{n\geq m_1-1}$ and $(z_n^{(2)})_{n\geq m_1-1}$ defined as $(z_n)_{n\geq m_1-1}$ above but using two different values $\tilde{\theta}^{(1)}$ and $\tilde{\theta}^{(2)}$, respectively. It is clear from (11) that for the choice $\tilde{\theta}^{(1)}:=\theta$ one has $b_n\leq z_n^{(1)}$ for all $n\geq m_1-1$. Moreover, since $b_{m_1+1}/b_{m_1}\leq z_{m_1+1}^{(1)}/z_{m_1}^{(1)}$, it holds by induction that

$$\begin{aligned}\frac{b_{n+1}}{b_n}&\leq 1+\frac{\eta_{n+1}}{\eta_n}(1-\eta_n)^3\left(1-\frac{b_{n-1}}{b_n}\right)+\eta_{n+1}\tilde{\theta}^2\\&\leq 1+\frac{\eta_{n+1}}{\eta_n}(1-\eta_n)^3\left(1-\frac{z_{n-1}^{(1)}}{z_n^{(1)}}\right)+\eta_{n+1}\tilde{\theta}^2=\frac{z_{n+1}^{(1)}}{z_n^{(1)}}\end{aligned}$$

also for all $n\geq m_1+1$. By a similar argument one shows that if $\tilde{\theta}^{(2)}:=[(1-\eta_{m_1})^2-1+\theta^2]^{1/2}$ then $b_n\geq z_n^{(2)}$ and $b_{n+1}/b_n\geq z_{n+1}^{(2)}/z_n^{(2)}$ for all $n\geq m_1-1$.

Let $\lambda' > 1$. Since $g_n^{(1)} \to \tilde{\theta}^{(1)}$ and $g_n^{(2)} \to \tilde{\theta}^{(2)}$ there is a $m_2 \geq m_1$ such that the following bounds apply

$$1 + \frac{\tilde{\theta}^{(2)}}{\lambda'}\sqrt{\eta_n} \leq \frac{z_n^{(2)}}{z_{n-1}^{(2)}} \qquad \text{and} \qquad \frac{z_n^{(1)}}{z_{n-1}^{(1)}} \leq 1 + \lambda'\tilde{\theta}^{(1)}\sqrt{\eta_n}$$

for all $n \geq m_2$. Consequently, for all $n \geq m_2$, we have that

$$\log\left(\frac{b_{n+k}}{b_n}\right) \leq \log\left(\frac{z_{n+k}^{(1)}}{z_n^{(1)}}\right) \leq \sum_{j=n+1}^{n+k} \log\left(1 + \lambda'\tilde{\theta}^{(1)}\sqrt{\eta_j}\right) \leq \lambda'\theta \sum_{j=n+1}^{n+k} \sqrt{\eta_n}.$$

Similarly, by the mean value theorem

$$\log\left(\frac{b_{n+k}}{b_n}\right) \geq \sum_{j=n+1}^{n+k} \log\left(1 + \frac{\tilde{\theta}^{(2)}}{\lambda'}\sqrt{\eta_j}\right) \geq \frac{\tilde{\theta}^{(2)}}{\lambda'(1+\lambda'^{-1}\tilde{\theta}^{(2)}\sqrt{\eta_n})} \sum_{j=n+1}^{n+k} \sqrt{\eta_j}$$

since η_n is decreasing. By letting the constant m_1 above be sufficiently large, the difference $|\tilde{\theta}^{(2)} - \theta|$ can be made arbitrarily small, and by increasing m_2, the constant $\lambda' > 1$ can be chosen arbitrarily close to one. □

3.3 Uniform target: path-wise behaviour

Section 3.2 characterised the behaviour of the sequence $\mathbb{E}\left[S_n\right]$ when the chain $(X_n)_{n\geq 2}$ follows the 'adaptive random walk' recursion (6). In this section, we shall verify that almost every sample path $(S_n)_{n\geq 1}$ of the same process are increasing.

Let us start by expressing the process S_n in terms of an auxiliary process $(Z_n)_{n\geq 1}$.

Lemma 13. *Let $u \in \mathbb{R}^d$ be a unit vector and suppose the process $(X_n, M_n, S_n)_{n\geq 1}$ is defined through* (3), (4) *and* (6)*, where $(W_n)_{n\geq 1}$ are i.i.d. following a spherically symmetric, non-degenerate distribution. Define the scalar process $(Z_n)_{n\geq 2}$ through*

$$Z_{n+1} := u^T \frac{X_{n+1} - M_n}{\|S_n^{1/2} u\|} \tag{13}$$

where $\|x\| := \sqrt{x^T x}$ stands for the Euclidean norm.

Then, the process $(Z_n, S_n)_{n\geq 2}$ follows

$$u^T S_{n+1} u = [1 + \eta_{n+1}(Z_{n+1}^2 - 1)]u^T S_n u \tag{14}$$

$$Z_{n+1} = \theta \tilde{W}_{n+1} + U_n Z_n \tag{15}$$

where $(\tilde{W}_n)_{n\geq 2}$ are non-degenerate i.i.d. random variables and $U_n := (1-\eta_n)\big(1+\eta_n(Z_n^2 - 1)\big)^{-1/2}$.

The proof of Lemma 13 is given in Appendix B.

It is immediate from (14) that only values $|Z_n| < 1$ can decrease $u^T S_n u$. On the other hand, if both η_n and $\eta_n Z_n^2$ are small, then the variable U_n is clearly close to unity. This suggests a nearly random walk behaviour of Z_n. Let us consider an auxiliary result quantifying the behaviour of this random walk.

Lemma 14. *Let $n_0 \geq 2$, suppose $\tilde{Z}_{n_0-1}$ is $\mathscr{F}_{n_0-1}$-measurable random variable and suppose $(\tilde{W}_n)_{n\geq n_0}$ are respectively $(\mathscr{F}_n)_{n\geq n_0}$-measurable and non-degenerate i.i.d. random variables. Define for $\tilde{Z}_n$ for $n \geq 2$ through*

$$\tilde{Z}_{n+1} = \tilde{Z}_n + \theta \tilde{W}_{n+1}.$$

Then, for any $N, \delta_1, \delta_2 > 0$, there is a $k_0 \geq 1$ such that

$$\mathbb{P}\left(\frac{1}{k} \sum_{j=1}^{k} \mathbb{1}_{\{|\tilde{Z}_{n+j}| \leq N\}} \geq \delta_1 \;\middle|\; \mathscr{F}_n \right) \leq \delta_2$$

a.s. for all $n \geq 1$ and $k \geq k_0$.

Proof. From the Kolmogorov-Rogozin inequality, Theorem 36 in Appendix C,

$$\mathbb{P}(\tilde{Z}_{n+j} - \tilde{Z}_n \in [x, x+2N] \mid \mathscr{F}_n) \leq c_1 j^{-1/2}$$

for any $x \in \mathbb{R}$, where the constant $c_1 > 0$ depends on N, θ and on the distribution of W_j. In particular, since $\tilde{Z}_{n+j} - \tilde{Z}_n$ is independent of $\tilde{Z}_n$, one may set $x = -Z_n - N$ above, and thus $\mathbb{P}\left(|\tilde{Z}_{n+j}| \leq N \;\middle|\; \mathscr{F}_n \right) \leq c_1 j^{-1/2}$. The estimate

$$\mathbb{E}\left[\frac{1}{k} \sum_{j=1}^{k} \mathbb{1}_{\{|\tilde{Z}_{n+j}| \leq N\}} \;\middle|\; \mathscr{F}_n \right] \leq \frac{c_1}{k} \sum_{j=1}^{k} j^{-1/2} \leq c_2 k^{-1/2}$$

implies $\mathbb{P}\big(k^{-1} \sum_{j=1}^{k} \mathbb{1}_{\{|\tilde{Z}_{n+j}| \leq N\}} \geq \delta_1 \;\big|\; \mathscr{F}_n\big) \leq \delta_1^{-1} c_2 k^{-1/2}$, concluding the proof. □

The technical estimate in the next Lemma 16 makes use of the above mentioned random walk approximation and guarantees ultimately a positive 'drift' for the eigenvalues of S_n. The result requires that the adaptation sequence $(\eta_n)_{n\geq 2}$ is 'smooth' in the sense that the quotients converge to one.

Assumption 15. The adaptation weight sequence $(\eta_n)_{n\geq 2} \subset (0,1)$ satisfies

$$\lim_{n\to\infty} \frac{\eta_{n+1}}{\eta_n} = 1.$$

Lemma 16. *Let $n_0 \geq 2$, suppose Z_{n_0-1} is $\mathscr{F}_{n_0-1}$-measurable, and assume $(Z_n)_{n\geq n_0}$ follows* (15) *with non-degenerate i.i.d. variables $(\tilde{W}_n)_{n\geq n_0}$ measurable with respect to $(\mathscr{F}_n)_{n\geq n_0}$, respectively, and the adaptation weights $(\eta_n)_{n\geq n_0}$ satisfy Assumption 15. Then, for any $C \geq 1$ and $\epsilon > 0$, there are indices $k \geq 1$ and $n_1 \geq n_0$ such that $\mathbb{P}\left(L_{n,k} \;\middle|\; \mathscr{F}_n \right) \leq \epsilon$ a.s. for all $n \geq n_1$, where*

$$L_{n,k} := \left\{ \sum_{j=1}^{k} \log\left[1 + \eta_{n+j}\left(Z_{n+j}^2 - 1\right)\right] < kC\eta_n \right\}.$$

Proof. Fix a $\gamma \in (0, 2/3)$. Define the sets $A_{n:j} := \cap_{i=n+1}^{j} \{Z_i^2 \le \eta_i^{-\gamma}\}$ and $A_i' := \{Z_i^2 > \eta_i^{-\gamma}\}$. Write the conditional expectation in parts as follows,

$$\begin{aligned}\mathbb{P}\left(L_{n,k} \mid \mathscr{F}_n\right) &= \mathbb{P}\left(L_{n,k}, A_{n:n+k} \mid \mathscr{F}_n\right) + \mathbb{P}\left(L_{n,k}, A_n' \mid \mathscr{F}_n\right) \\ &\quad + \sum_{i=n+1}^{n+k} \mathbb{P}\left(L_{n,k}, A_{n:i-1}, A_i' \mid \mathscr{F}_n\right).\end{aligned} \tag{16}$$

Let $\omega \in A_i'$ for any $n < i \le n+k$ and compute

$$\begin{aligned}\log\left[1+\eta_i(Z_i^2-1)\right] &\ge \log\left[1+\eta_i(\eta_i^{-\gamma}-1)\right] \ge \log\left[1+2\eta_i kC\right] \\ &\ge \frac{2\eta_i kC}{1+2\eta_i kC} \ge kC\eta_n\end{aligned}$$

whenever $n \ge n_0$ is sufficiently large, since $\eta_n \to 0$, and by Assumption 15. That is, if n is sufficiently large, all but the first term in the right hand side of (16) are a.s. zero. It remains to show the inequality for the first.

Suppose now that $Z_n^2 \le \eta_n^{-\gamma}$. One may estimate

$$\begin{aligned}U_n &= (1-\eta_n)^{1/2}\left(1-\frac{\eta_n Z_n^2}{1-\eta_n+\eta_n Z_n^2}\right)^{1/2} \\ &\ge (1-\eta_n)^{1/2}\left(1-\frac{\eta_n^{1-\gamma}}{1-\eta_n}\right)^{1/2} \\ &\ge (1-\eta_n^{1-\gamma})^{1/2}\left(\frac{1-2\eta_n^{1-\gamma}}{1-\eta_n}\right)^{1/2} \ge 1-c_1\eta_n^{1-\gamma}\end{aligned}$$

where $c_1 := 2\sup_{n\ge n_0}(1-\eta_n)^{-1/2} < \infty$. Observe also that $U_n \le 1$.

Let $k_0 \ge 1$ be from Lemma 14 applied with $N = \sqrt{8C+1}+1$, $\delta_1 = 1/8$ and $\delta_2 = \epsilon$, and fix $k \ge k_0+1$. Let $n \ge n_0$ and define an auxiliary process $(\tilde{Z}_j^{(n)})_{j\ge n_0-1}$ as $\tilde{Z}_j^{(n)} \equiv Z_j$ for $n_0-1 \le j \le n+1$, and for $j > n+1$ through

$$\tilde{Z}_j^{(n)} = Z_{n+1} + \theta \sum_{i=n+2}^{j} \tilde{W}_i.$$

For any $n+2 \le j \le n+k$ and $\omega \in A_{n:j}$, the difference of $\tilde{Z}_j^{(n)}$ and Z_j can be bounded by

$$\begin{aligned}|\tilde{Z}_{j+1}^{(n)} - Z_{j+1}| &\le |Z_j||1-U_j| + |\tilde{Z}_j^{(n)} - Z_j| \le c_1\eta_j^{1-\frac{3}{2}\gamma} + |\tilde{Z}_j^{(n)} - Z_j| \le \cdots \\ &\le c_1\sum_{i=n+1}^{j}\eta_i^{1-\frac{3}{2}\gamma} \le c_1\eta_n^{1-\frac{3}{2}\gamma}\sum_{i=n+1}^{j}\left(\frac{\eta_i}{\eta_n}\right)^{1-\frac{3}{2}\gamma} \le c_2(j-n)\eta_n^{1-\frac{3}{2}\gamma}\end{aligned}$$

by Assumption 15. Therefore, for sufficiently large $n \ge n_0$, the inequality $|\tilde{Z}_j^{(n)} - Z_j| \le 1$ holds for all $n \le j \le n+k$ and $\omega \in A_{n:n+k}$. Now, if $\omega \in A_{n:n+k}$, the following bound holds

$$\begin{aligned}\log\left[1+\eta_j(Z_j^2-1)\right] &\ge \log\left[1+\eta_j(\min\{N,|Z_j|\}^2-1)\right] \\ &\ge \mathbb{1}_{\{|\tilde{Z}_j^{(n)}|>N\}}\log\left[1+\eta_j((N-1)^2-1)\right] + \mathbb{1}_{\{|\tilde{Z}_j^{(n)}|\le N\}}\log\left[1-\eta_j\right] \\ &\ge \mathbb{1}_{\{|\tilde{Z}_j^{(n)}|>N\}}(1-\beta_j)\eta_j 8C - \mathbb{1}_{\{|\tilde{Z}_j^{(n)}|\le N\}}(1+\beta_j)\eta_j\end{aligned}$$

by the mean value theorem, where the constant $\beta_j = \beta_j(C,\eta_j) \in (0,1)$ can be selected arbitrarily small whenever j is sufficiently large. Using this estimate, one can write for $\omega \in A_{n:n+k}$

$$\sum_{j=1}^{k} \log\left[1+\eta_{n+j}\left(Z_{n+j}^2-1\right)\right] \geq (1-\beta_n)\sum_{j\in I_{n+1:k}^+} \eta_{n+j} 8C - (1+\beta_n)\sum_{j=1}^{k}\eta_{n+j}$$

where $I_{n+1:k}^+ := \{j \in [1,k] : \tilde{Z}_{n+j}^{(n)} > N\}$. Define the sets

$$B_{n,k} := \left\{ \frac{1}{k-1}\sum_{j=1}^{k-1} \mathbb{1}_{\{|\tilde{Z}_{n+j+1}|\leq N\}} \leq \delta_1 \right\}.$$

Within $B_{n,k}$, it clearly holds that $\#I_{n+1:k}^+ \geq k-1-(k-1)\delta_1 = 7(k-1)/8$. Thereby, for all $\omega \in B_{n,k}\cap A_{n:n+k}$

$$\begin{aligned}\sum_{j=1}^{k} &\log\left[1+\eta_{n+j}\left(Z_{n+j}^2-1\right)\right] \\ &\geq \eta_n k\left[(1-\beta_n)\frac{7}{2}\left(\inf_{1\leq j\leq k}\frac{\eta_{n+j}}{\eta_n}\right)C - (1+\beta_n)\left(\sup_{1\leq j\leq k}\frac{\eta_{n+j}}{\eta_n}\right)\right] \geq kC\eta_n\end{aligned}$$

for sufficiently large $n \geq 1$, as then the constant β_n can be chosen small enough, and by Assumption 15. In other words, if $n \geq 1$ is sufficiently large, then $B_{n,k}\cap A_{n:n+k}\cap L_{n,k} = \emptyset$. Now, Lemma 14 yields

$$\begin{aligned}\mathbb{P}\left(L_{n,k}, A_{n:n+k} \,\middle|\, \mathscr{F}_n\right) &= \mathbb{P}\left(L_{n,k}, A_{n:n+k}, B_{n,k} \,\middle|\, \mathscr{F}_n\right) \\ &\quad + \mathbb{P}\left(L_{n,k}, A_{n:n+k}, B_{n,k}^{\complement} \,\middle|\, \mathscr{F}_n\right) \\ &\leq \mathbb{P}\left(B_{n,k}^{\complement} \mid \mathscr{F}_n\right) \leq \epsilon.\end{aligned}$$

□

Using the estimate of Lemma 16, it is relatively easy to show that $u^T S_n u$ tends to infinity, if the adaptation weights satisfy an additional assumption.

Assumption 17. The adaptation weight sequence $(\eta_n)_{n\geq 2} \subset (0,1)$ is in ℓ^2 but not in ℓ^1, that is,

$$\sum_{n=2}^{\infty}\eta_n = \infty \quad \text{and} \quad \sum_{n=2}^{\infty}\eta_n^2 < \infty.$$

Theorem 18. *Assume that $(X_n)_{n\geq 2}$ follows the 'adaptive random walk' recursion* (6) *and the adaptation weights $(\eta_n)_{n\geq 2}$ satisfy Assumptions 15 and 17. Then, for any unit vector $u \in \mathbb{R}^d$, the process $u^T S_n u \to \infty$ almost surely.*

Proof. The proof is based on the estimate of Lemma 16 applied with a similar martingale argument as in Vihola (2009).

Let $k \geq 2$ be from Lemma 16 applied with $C = 4$ and $\epsilon = 1/2$. Denote $\ell_i := ki + 1$ for $i \geq 0$ and, inspired by (14), define the random variables $(T_i)_{i\geq 1}$ by

$$T_i := \min\left\{ kM\eta_{\ell_{i-1}}, \sum_{j=\ell_{i-1}+1}^{\ell_i} \log\left[1+\eta_j\left(Z_j^2-1\right)\right] \right\}$$

with the convention that $\eta_0 = 1$. Form a martingale $(Y_i, \mathscr{G}_i)_{i\geq 1}$ with $Y_1 \equiv 0$ and having differences $\mathrm{d}Y_i := T_i - \mathbb{E}\left[T_i \,\middle|\, \mathscr{G}_{i-1} \right]$ and where $\mathscr{G}_1 \equiv \{\emptyset, \Omega\}$ and $\mathscr{G}_i := \mathscr{F}_{\ell_i}$ for $i \geq 1$. By Assumption 17,

$$\sum_{i=2}^{\infty} \mathbb{E}\left[\mathrm{d}Y_i^2 \right] \leq c \sum_{i=1}^{\infty} \eta_{\ell_i}^2 < \infty$$

with a constant $c = c(k, C) > 0$, so Y_i is a L^2-martingale and converges a.s. to a finite limit M_∞ (e.g. Hall and Heyde 1980, Theorem 2.15).

By Lemma 16, the conditional expectation satisfies

$$\mathbb{E}\left[T_{i+1} \,\middle|\, \mathscr{G}_i \right] \geq kC\eta_{\ell_i}(1-\epsilon) + \sum_{j=\ell_i+1}^{\ell_{i+1}} \log(1-\eta_j)\epsilon \geq k\eta_{\ell_i}$$

when i is large enough, and where the second inequality is due to Assumption 15. This implies, with Assumption 17, that $\sum_i \mathbb{E}\left[T_i \,\middle|\, \mathscr{G}_{i-1} \right] = \infty$ a.s., and since Y_i converges a.s. to a finite limit, it holds that $\sum_i T_i = \infty$ a.s.

By (14), one may estimate for any $n = \ell_m$ with $m \geq 1$ that

$$\log(u^T S_n u) \geq \log(u^T S_1 u) + \sum_{i=1}^{m} T_i \to \infty$$

as $m \to \infty$. Simple deterministic estimates conclude the proof for the intermediate values of n. □

3.4 Stability with one-dimensional uniformly continuous log-density

In this section, the above analysis of the 'adaptive random walk' is extended to imply that $\liminf_{n\to\infty} S_n > 0$ for the one-dimensional AM algorithm, assuming $\log \pi$ uniformly continuous. The result follows similarly as in Theorem 18, by coupling the AM process with the 'adaptive random walk' whenever S_n is small enough to ensure that the acceptance probability is sufficiently close to one.

Theorem 19. *Assume $d = 1$ and $\log \pi$ is uniformly continuous, and that the adaptation weights $(\eta_n)_{n\geq 2}$ satisfy Assumptions 15 and 17. Then, there is a constant $b > 0$ such that $\liminf_{n\to\infty} S_n \geq b$.*

Proof. Fix a $\delta \in (0, 1)$. Due to the uniform continuity of $\log \pi$, there is a $\tilde{\delta} > 0$ such that

$$\log \pi(y) - \log \pi(x) \geq \frac{1}{2} \log\left(1 - \frac{\delta}{2}\right)$$

for all $|x - y| \leq \tilde{\delta}_1$. Choose $\tilde{M} > 0$ sufficiently large so that $\int_{\{|z|\leq \tilde{M}\}} \tilde{q}(z)\mathrm{d}z \geq \sqrt{1-\delta/2}$. Denote by

$$Q_q(x,A) := \int_A q(y-x)\mathrm{d}y$$

the random walk transition kernel with increment distribution q, and observe that the 'adaptive random walk' recursion (6) can be written as "$X_{n+1} \sim Q_{\tilde{q}_{S_n}}(X_n, \cdot)$." For any $x \in \mathbb{R}^d$ and measurable $A \subset \mathbb{R}^d$

$$|Q_{\tilde{q}_s}(x,A) - P_{\tilde{q}_s}(x,A)| \leq 2\left[1 - \int \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} \tilde{q}_s(y-x)\mathrm{d}y\right]$$
$$\leq 2\left[1 - \int_{\{|z|\leq \tilde{M}\}} \min\left\{1, \frac{\pi(x+\sqrt{\theta s}z)}{\pi(x)}\right\} \tilde{q}(z)\mathrm{d}z\right].$$

Now, $|Q_{\tilde{q}_s}(x,A) - P_{\tilde{q}_s}(x,A)| \leq \delta$ whenever $\sqrt{\theta s}z \leq \tilde{\delta_1}$ for all $|z| \leq \tilde{M}$. In other words, there exists a $\mu = \mu(\delta) > 0$ such that whenever $s < \mu$, the total variation norm $\|Q_{\tilde{q}_s}(x,\cdot) - P_{\tilde{q}_s}(x,\cdot)\| \leq \delta$.

Next we shall consider a 'adaptive random walk' process to be coupled with $(X_n, M_n, S_n)_{n\geq 1}$. Let $n,k \geq 1$ and define the random variables $(\tilde{X}_j^{(n)}, \tilde{M}_j^{(n)}, \tilde{S}_j^{(n)})_{j\in[n,n+k]}$ by setting $(\tilde{X}_n^{(n)}, \tilde{M}_n^{(n)}, \tilde{S}_n^{(n)}) \equiv (X_n, M_n, S_n)$ and

$$\begin{aligned}
\tilde{X}_{j+1}^{(n)} &\sim Q_{\tilde{q}_{\tilde{S}_j^{(n)}}}(\tilde{X}_j^{(n)}, \cdot), \\
\tilde{M}_{j+1}^{(n)} &:= (1-\eta_{j+1})\tilde{M}_j^{(n)} + \eta_{j+1}\tilde{X}_{j+1}^{(n)} \quad \text{and} \\
\tilde{S}_{j+1}^{(n)} &:= (1-\eta_{j+1})\tilde{S}_j^{(n)} + \eta_{j+1}(\tilde{X}_{j+1}^{(n)} - \tilde{M}_j^{(n)})^2
\end{aligned}$$

for $j+1 \in [n+1, n+k]$. The variable $\tilde{X}_{n+1}^{(n)}$ can be selected so that $\mathbb{P}(\tilde{X}_{n+1}^{(n)} = X_{n+1} \mid \mathscr{F}_n) = 1 - \|P_{\tilde{q}_{S_n}}(X_n, \cdot) - Q_{\tilde{q}_{\tilde{S}_n^{(n)}}}(\tilde{X}_n^{(n)}, \cdot)\|$; see Theorem 37 in Appendix D. Consequently, $\mathbb{P}(\tilde{X}_{n+1}^{(n)} \neq X_{n+1}, S_n < \mu \mid \mathscr{F}_n) \leq \delta$. By the same argument, $\tilde{X}_{n+2}^{(n)}$ can be chosen so that

$$\mathbb{P}\big(\tilde{X}_{n+2}^{(n)} \neq X_{n+2}, \tilde{X}_{n+1}^{(n)} = X_{n+1}, S_{n+1} < \mu \,\big|\, \sigma(\mathscr{F}_{n+1}, \tilde{X}_{n+1}^{(n)})\big) \leq \delta$$

since if $\tilde{X}_{n+1}^{(n)} = X_{n+1}$, then also $\tilde{S}_{n+1}^{(n)} = S_{n+1}$. This implies

$$\mathbb{P}\big(\{\tilde{X}_{n+2}^{(n)} \neq X_{n+2}\} \cup \{\tilde{X}_{n+1}^{(n)} \neq X_{n+1}\} \cap B_{n:n+2} \,\big|\, \mathscr{F}_n\big) \leq 2\delta$$

where $B_{n:j} := \cap_{i=n}^{j-1}\{S_i < \mu\}$ for $j > n$. The same argument can be repeated to construct $(\tilde{X}_j^{(n)})_{j\in[n,n+k]}$ so that

$$\mathbb{P}\left(D_{n:n+k} \,\middle|\, \mathscr{F}_n\right) \geq 1 - k\delta \tag{17}$$

where $D_{n:n+k} := \bigcap_{j=n}^{n+k}\{\tilde{X}_j^{(n)} = X_j\} \cup B_{n:n+k}^{\complement}$.

Apply Lemma 16 with $C = 18$ and $\epsilon = 1/6$ to obtain $k \geq 1$, and fix $\delta = \epsilon/k$. Denote $\ell_i := ik+1$ for any $i \geq 0$, and define the random variables $(T_i)_{i\geq 1}$ by

$$T_i := \mathbb{1}_{\{S_{\ell_{i-1}} < \mu/2\}} \min\left\{kM\eta_{\ell_{i-1}}, \sum_{j=\ell_{i-1}+1}^{\ell_i} \log\left[1 + \eta_j\left(Z_j^2 - 1\right)\right]\right\} \tag{18}$$

where Z_j are defined as (13).

Define also $\tilde{T}_i$ similarly as T_i, but having $\tilde{Z}_j^{(\ell_{i-1})}$ with $j \in [\ell_{i-1}+1, \ell_i]$ in the right hand side of (18), defined as $\tilde{Z}_{\ell_{i-1}}^{(\ell_{i-1})} \equiv Z_{\ell_{i-1}}$ and by

$$\tilde{Z}_j^{(\ell_{i-1})} := \left(\tilde{X}_j^{(\ell_{i-1})} - \tilde{M}_{j-1}^{(\ell_{i-1})}\right) \Big/ \sqrt{\tilde{S}_{j-1}^{(\ell_{i-1})}}.$$

for $j \in [\ell_{i-1}+1, \ell_i]$. Notice that T_i coincides with $\tilde{T}_i$ in $B_{\ell_{i-1}:\ell_i} \cap D_{\ell_{i-1}:\ell_i}$. Observe also that $\tilde{X}_j^{(\ell_{i-1})}$ follows the 'adaptive random walk' equation (6) for $j \in [\ell_{i-1}+1, \ell_i]$, and hence $\tilde{Z}_j^{(\ell_{i-1})}$ follows (15). Consequently, denoting $\mathscr{G}_i := F_{\ell_i}$, Lemma 16 guarantees that

$$\mathbb{P}\left(L_{\ell_{i-1},k} \,\middle|\, \mathscr{G}_i \right) \le \epsilon \tag{19}$$

where $L_{\ell_{i-1},k} := \{\tilde{T}_i < kM\eta_{\ell_{i-1}}\}$.

Let us show next that whenever $S_{\ell_{i-1}}$ is small, the variable T_i is expected to have a positive value proportional to the adaptation weight,

$$\mathbb{E}\left[T_i \,\middle|\, \mathscr{G}_{i-1} \right] \mathbb{1}_{\{S_{\ell_{i-1}} < \mu/2\}} \ge k\eta_{\ell_{i-1}} \mathbb{1}_{\{S_{\ell_{i-1}} < \mu/2\}} \tag{20}$$

almost surely for any sufficiently large $i \ge 1$. Write first

$$\begin{aligned}\mathbb{E}\left[T_i \,\middle|\, \mathscr{G}_{i-1} \right] \mathbb{1}_{\{S_{\ell_{i-1}} < \mu/2\}} &= \mathbb{E}\left[(\mathbb{1}_{B^{\complement}_{\ell_{i-1}:\ell_i}} + \mathbb{1}_{B_{\ell_{i-1}:\ell_i}}) T_i \,\middle|\, \mathscr{G}_{i-1} \right] \mathbb{1}_{\{S_{\ell_{i-1}} < \mu/2\}} \\ &\ge \mathbb{E}\left[\mathbb{1}_{B^{\complement}_{\ell_{i-1}:\ell_i}} \min\left\{ kC\eta_{\ell_{i-1}}, \frac{\mu}{2} + \xi_i \right\} + \mathbb{1}_{B_{\ell_{i-1}:\ell_i}} \xi_i \,\middle|\, \mathscr{G}_{i-1} \right] \mathbb{1}_{\{S_{\ell_{i-1}} < \mu/2\}}\end{aligned}$$

where the lower bound ξ_i of T_i is given as

$$\xi_i := \sum_{j=\ell_{i-1}+1}^{\ell_i} \log(1-\eta_j).$$

By Assumption 15, $\xi_i \ge -2k\eta_{\ell_{i-1}} \ge -\mu/4$ for any sufficiently large i. Therefore, whenever $\mathbb{P}\left(B^{\complement}_{\ell_{i-1}:\ell_i} \,\middle|\, \mathscr{G}_{i-1} \right) \ge \epsilon = 3/C$, it holds that

$$\mathbb{E}\left[T_i \,\middle|\, \mathscr{G}_{i-1} \right] \mathbb{1}_{\{S_{\ell_{i-1}} < \mu/2\}} \ge k\eta_{\ell_{i-1}} \mathbb{1}_{\{S_{\ell_{i-1}} < \mu/2\}}$$

for any sufficiently large i. On the other hand, if $\mathbb{P}\left(B^{\complement}_{\ell_{i-1}:\ell_i} \,\middle|\, \mathscr{G}_{i-1} \right) \le \epsilon$, then by defining

$$E_i := B^{\complement}_{\ell_{i-1}:\ell_i} \cup D^{\complement}_{\ell_{i-1}:\ell_i} \cup L_{\ell_{i-1},k}$$

one has by (17) and (19) that $\mathbb{P}(E_i) \le 3\epsilon$, and consequently

$$\begin{aligned}\mathbb{E}\left[T_i \,\middle|\, \mathscr{G}_{i-1} \right] &\ge \mathbb{P}\left(E_i^{\complement} \,\middle|\, \mathscr{G}_{i-1} \right) \xi_i + \mathbb{E}\left[\mathbb{1}_{E_i} \tilde{T}_i \,\middle|\, \mathscr{G}_{i-1} \right] \\ &\ge 3\epsilon\xi_i + (1-3\epsilon)kC\eta_{\ell_{i-1}} \ge k\eta_{\ell_{i-1}}.\end{aligned}$$

This establishes (20).

Define the stopping times $\tau_1 \equiv 1$ and for $n \geq 2$ through $\tau_n := \inf\{i > \tau_{n-1} : S_{\ell_{i-1}} \geq \mu/2, S_{\ell_i} < \mu/2\}$ with the convention that $\inf \emptyset = \infty$. That is, τ_i record the times when S_{ℓ_i} enters $(0, \mu/2]$. Using τ_i, define the latest such time up to n by $\sigma_n := \sup\{\tau_i : i \geq 1, \tau_i \leq n\}$. As in Theorem 18, define the almost surely converging martingale $(Y_i, \mathscr{G}_i)_{i\geq 1}$ with $Y_1 \equiv 0$ and having the differences $\mathrm{d}Y_i := (T_i - \mathbb{E}\left[T_i \mid \mathscr{G}_{i-1}\right])$ for $i \geq 2$.

It is sufficient to show that $\liminf_{i\to\infty} S_{\ell_i} \geq b := \mu/4 > 0$ almost surely. If there is a finite $i_0 \geq 1$ such that $S_{\ell_i} \geq \mu/2$ for all $i \geq i_0$, the claim is trivial. Let us consider for the rest of the proof the case that $\{S_{\ell_i} < \mu/2\}$ happens for infinitely many indices $i \geq 1$.

For any $m \geq 2$ such that $S_{\ell_m} < \mu/2$, one can write

$$\begin{aligned} \log S_{\ell_m} &\geq \log S_{\ell_{\sigma_m}} + \sum_{i=\sigma_m+1}^{m} T_i \\ &\geq \log S_{\ell_{\sigma_m}} + (Y_m - Y_{\sigma_m}) + \sum_{i=\sigma_m+1}^{m} k\eta_{\ell_{i-1}} \end{aligned} \tag{21}$$

since then $S_{\ell_i} < \mu/2$ for all $i \in [\sigma_m, m-1]$ and hence also $\mathbb{E}\left[T_i \mid \mathscr{G}_{i-1}\right] \geq k\eta_{\ell_{i-1}}$.

Suppose for a moment that there is a positive probability that S_{ℓ_m} stays within $(0, \mu/2)$ indefinitely, starting from some index $m_1 \geq 1$. Then, there is an infinite τ_i and consequently $\sigma_m \leq \sigma < \infty$ for all $m \geq 1$. But as Y_m converges, $|Y_m - Y_{\sigma_m}|$ is a.s. finite, and since $\sum_m \eta_{\ell_m} = \infty$ by Assumptions 15 and 17, the inequality (21) implies that $S_{\ell_m} \geq \mu/2$ for sufficiently large m, which is a contradiction. That is, the stopping times τ_i for all $i \geq 1$ must be a.s. finite, whenever $S_{\ell_m} < \mu/2$ for infinitely many indices $m \geq 1$.

For the rest of the proof, suppose $S_{\ell_m} < \mu/2$ for infinitely many indices $m \geq 1$. Observe that since $Y_m \to Y_\infty$, there exists an a.s. finite index m_2 so that $Y_m - Y_\infty \geq -1/2 \log 2$ for all $m \geq m_2$. As $\eta_n \to 0$ and $\sigma_m \to \infty$, there is an a.s. finite m_3 such that $\xi_{\sigma_{m-1}} \geq -1/2 \log 2$ for all $m \geq m_3$. For all $m \geq \max\{m_2, m_3\}$ and whenever $S_{\ell_m} < \mu/2$, it thereby holds that

$$\begin{aligned} \log S_{\ell_m} &\geq \log S_{\ell_{\sigma_m}} - (Y_m - Y_{\sigma_m}) \geq \log S_{\ell_{\sigma_{m-1}}} + \xi_{\sigma_m} - \frac{1}{2}\log 2 \\ &\geq \log \frac{\mu}{2} - \log 2 = \log b. \end{aligned}$$

The case $S_{\ell_m} \geq \mu/2$ trivially satisfies the above estimate, concluding the proof. □

As a consequence of Theorem 19, one can establish a strong law of large numbers for the unconstrained AM algorithm running with a Laplace target distribution. Essentially, the only ingredient that needs to be checked is that the simultaneous geometric ergodicity condition holds. This is verified in the next lemma, whose proof is given in Appendix E.

Lemma 20. *Suppose that the template proposal distribution $\tilde{q}$ is everywhere positive and non-increasing away from the origin: $\tilde{q}(z) \geq \tilde{q}(w)$ for all $|z| \leq |w|$. Suppose also that $\pi(x) := \frac{1}{2b}\exp\left(-\frac{|x-m|}{b}\right)$ with a mean $m \in \mathbb{R}$ and a scale $b > 0$. Then, for all $L > 0$, there are positive* *constants M, b such that the following drift and minorisation condition are satisfied for all $s \geq L$ and measurable $A \subset \mathbb{R}$*

$$P_s V(x) \leq \lambda_s V(x) + b\mathbb{1}_C(x), \qquad \forall x \in \mathbb{R} \tag{22}$$
$$P_s(x, A) \geq \delta_s \nu(A), \qquad \forall x \in C \tag{23}$$

where $V : \mathbb{R} \to [1, \infty)$ is defined as $V(x) := (\sup_z \pi(z))^{1/2}\pi^{-1/2}(x)$, the set $C := [m-M, m+M]$, the probability measure μ is concentrated on C and $P_s V(x) := \int V(y)P_s(x, \mathrm{d}y)$. Moreover, $\lambda_s, \delta_s \in (0,1)$ satisfy for all $s \geq L$

$$\max\{(1-\lambda_s)^{-1}, \delta_s^{-1}\} \leq cs^{\gamma} \tag{24}$$

for some constants $c, \gamma > 0$ that may depend on L.

Theorem 21. *Assume the adaptation weights $(\eta_n)_{n\geq 2}$ satisfy Assumptions 15 and 17, and the template proposal density $\tilde{q}$ and the target distribution π satisfy the assumptions in Lemma 20. If the functional f satisfies $\sup_{x\in\mathbb{R}} \pi^{-\gamma}(x)|f(x)| < \infty$ for some $\gamma \in (0, 1/2)$. Then, $n^{-1}\sum_{k=1}^{n} f(X_k) \to \int f(x)\pi(x)\mathrm{d}x$ almost surely as $n \to \infty$.*

Proof. The conditions of 19 are clearly satisfied implying that for any $\epsilon > 0$ there is a $\kappa = \kappa(\epsilon) > 0$ such that the event

$$B_\kappa := \left\{\inf_n S_n \geq \kappa\right\}$$

has a probability $\mathbb{P}(B_\kappa) \geq 1 - \epsilon$.

The inequalities (22) and (23) of Lemma 20 with the bound (24) imply, using (Saksman and Vihola 2010, Proposition 7 and Lemma 12), that for any $\beta > 0$ there is a constant $A = A(\kappa, \epsilon, \beta) < \infty$ such that $\mathbb{P}(B_\kappa \cap \{\max\{|S_n|, |M_n|\} > An^\beta\}) \leq \epsilon$. Let us define the sequence of truncation sets

$$K_n := \{(m, s) \in \mathbb{R} \times \mathbb{R}_+ : \lambda_{\min}(s) \geq \kappa, \max\{|s|, |m|\} \leq An^\beta\}$$

for $n \geq 1$. Construct an auxiliary truncated process $(\tilde{X}_n, \tilde{M}_n, \tilde{S}_n)_{n\geq 1}$, starting from $(\tilde{X}_1, \tilde{M}_1, \tilde{S}_1) \equiv (X_1, M_1, S_1)$ and for $n \geq 2$ through

$$\begin{aligned}
\tilde{X}_{n+1} &\sim P_{\tilde{q}_{\tilde{S}_n}}(\tilde{X}_n, \cdot) \\
(\tilde{M}_{n+1}, \tilde{S}_{n+1}) &= \sigma_{n+1}\Big[(\tilde{M}_n, \tilde{S}_n),\, \eta_{n+1}\big(\tilde{X}_{n+1} - \tilde{M}_n, (\tilde{X}_{n+1} - \tilde{M}_n)^2 - \tilde{S}_n\big)\Big]
\end{aligned}$$

where the truncation function $\sigma_{n+1} : (K_n) \times (\mathbb{R} \times \mathbb{R}) \to K_n$ is defined as

$$\sigma_{n+1}(z, z') = \begin{cases} z + z', & \text{if } z + z' \in K_n \\ z, & \text{otherwise.} \end{cases}$$

Observe that this constrained process coincides with the AM process with probability $\mathbb{P}\big(\forall n \geq 1 : (\tilde{X}_n, \tilde{M}_n, \tilde{S}_n) = (X_n, M_n, S_n)\big) \geq 1 - 2\epsilon$. Moreover, (Saksman and Vihola 2010, Theorem 2) implies that a strong law of large numbers holds for the truncated process $(\tilde{X}_n)_{n\geq 1}$, since $\sup_x |f(x)|V^{-\alpha}(x) < \infty$ for some $\alpha \in (0, 1 - \beta)$, by selecting $\beta > 0$ above sufficiently small. Since $\epsilon > 0$ was arbitrary, the strong law of large numbers holds for $(X_n)_{n\geq 1}$. □

4 AM with a fixed proposal component

This section deals with the modification due to Roberts and Rosenthal Roberts and Rosenthal (2009), including a fixed component in the proposal distribution. In terms of Section 2, the mixing parameter in (5) satisfies $0 < \beta < 1$. Theorem 24 shows that the fixed proposal component guarantees, with a verifiable non-restrictive Assumption 22, that the eigenvalues of the adapted covariance parameter S_n are bounded away from zero. As in Section 3.4, this result implies an ergodicity result, Theorem 29.

Let us start by formulating the key assumption that, intuitively speaking, assures that the adaptive chain $(X_n)_{n\geq 1}$ will have 'uniform mobility' regardless of the adaptation parameter $s \in \mathscr{C}^d$.

Assumption 22. There exist a compactly supported probability measure ν that is absolutely continuous with respect to the Lebesgue measure, constants $\delta > 0$ and $c < \infty$ and a measurable mapping $\xi : \mathbb{R}^d \times \mathscr{C}^d \to \mathbb{R}^d$ such that for all $x \in \mathbb{R}^d$ and $s \in \mathscr{C}^d$,

$$\|\xi(x,s) - x\| \leq c \qquad \text{and} \qquad P_{q_s}(x,A) \geq \delta\nu\big(A - \xi(x,s)\big)$$

for all measurable sets $A \subset \mathbb{R}^d$, where $A - y := \{x - y : x \in A\}$ is the translation of the set A by $y \in \mathbb{R}^d$.

Remark 23. In the case of the AM algorithm with a fixed proposal component, one is primarily interested in the case where $\xi(x,s) = \xi(x)$ and for all $x \in \mathbb{R}^d$

$$\beta q_{\text{fix}}(x-y)\min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} \geq \delta v\big(y - \xi(x)\big)$$

for all $y \in \mathbb{R}^d$, where v is a uniform density on some ball. Then, since $P_{q_s} = (1-\beta)P_{\tilde{q}_s} + \beta P_{q_{\text{fix}}}$,

$$P_{q_s}(x,A) \geq \beta P_{q_{\text{fix}}}(x,A) \geq \delta \int_A v(y - \xi)\mathrm{d}y$$

and Assumption 22 is fulfilled by the measure $\nu(A) := \int_A v(y)\mathrm{d}y$.

Having Assumption 22, the lower bound on the eigenvalues of S_n can be obtained relatively easily, by a martingale argument similar to the one used in Section 3 and in Vihola (2009).

Theorem 24. *Let $(X_n, M_n, S_n)_{n\geq 1}$ be an AM process as defined in Section 2 satisfying Assumption 22. Moreover, suppose that the adaptation weights $(\eta_n)_{n\geq 2}$ satisfy Assumptions 15 and 17. Then,*

$$\liminf_{n\to\infty} \inf_{w\in\mathscr{S}^d} w^T S_n w > 0$$

where $\mathscr{S}^d$ stands for the unit sphere.

Proof. Let us first introduce independent binary auxiliary variables $(Z_n)_{n\geq 2}$ with $Z_1 \equiv 0$, and through

$$\begin{aligned}\mathbb{P}\left(Z_{n+1} = 1 \,\middle|\, X_n, M_n, S_n, Z_n\right) &= \delta \\ \mathbb{P}\left(Z_{n+1} = 0 \,\middle|\, X_n, M_n, S_n, Z_n\right) &= (1-\delta).\end{aligned}$$

Using this auxiliary variable, we can assume X_n to follow[3]

$$X_{n+1} = Z_{n+1}(U_{n+1} + \Xi_n) + (1 - Z_{n+1})R_{n+1}$$

where $U_{n+1} \sim \nu(\cdot)$ is independent of $\mathscr{F}_n$ and Z_{n+1}, the random variable $\Xi_n := \xi(X_n, S_n)$ is $\mathscr{F}_n$-measurable, and R_{n+1} is distributed according to the 'residual' transition kernel $\check{P}_{S_n}(X_n, A) := (1-\delta)^{-1}[P_{q_{S_n}}(X_n, A) - \delta\nu(A - \Xi_n)]$, valid by Assumption 22.

Define $\mathscr{S}(w,\gamma) := \{v \in \mathscr{S}^d : \|w - v\| \le \gamma\}$, the segment of the unit sphere centred at $w \in \mathscr{S}^d$ and having the radius $\gamma > 0$. Fix a unit vector $w \in \mathscr{S}^d$ and define the following random variables

$$\Gamma_{n+2}^{(\gamma)} := \inf_{v \in \mathscr{S}(w,\gamma)} \left(|v^T(X_{n+1} - M_n)|^2 + |v^T(X_{n+2} - M_{n+1})|^2 \right)$$

for all $n \ge 1$. Denote $G_{n+1} := X_{n+1} - M_n$ and $E_{n+1} := \Xi_{n+1} - X_{n+1}$, and observe that whenever $Z_{n+2} = 1$, it holds that

$$\begin{aligned} X_{n+2} - M_{n+1} &= U_{n+2} + X_{n+1} - M_{n+1} + E_{n+1} \\ &= U_{n+2} + (1 - \eta_{n+1})G_{n+1} + E_{n+1} \end{aligned}$$

and we may write

$$Z_{n+2}\Gamma_{n+2}^{(\gamma)} = Z_{n+2} \inf_{v \in \mathscr{S}(w,\gamma)} \left(|v^T G_{n+1}|^2 + |v^T(U_{n+2} + \lambda_{n+1}G_{n+1} + E_{n+1})|^2 \right)$$

where $\lambda_n := 1 - \eta_n \in (0,1)$ for all $n \ge 2$. Consequently, we may apply Lemma 25 below to find constants $\gamma, \mu > 0$ such that

$$\mathbb{P}\left(Z_{n+2}\Gamma_{n+2}^{(\gamma)} \ge \mu \;\middle|\; \mathscr{F}_n \right) \ge \frac{\delta}{2}. \tag{25}$$

Hereafter, assume $\gamma > 0$ is fixed such that (25) holds, and denote $\Gamma_{n+2} := \Gamma_{n+2}^{(\gamma)}$ and $\mathscr{S}(w) := \mathscr{S}(w,\gamma)$.

Consider the random variables

$$\begin{aligned} D_{n+2} := \inf_{v \in \mathscr{S}(w)} \left(\eta_{n+1}|v^T(X_{n+1} - M_n)|^2 + \eta_{n+2}|v^T(X_{n+2} - M_{n+1})|^2 \right) \\ \ge \min\{\eta_{n+1}, \eta_{n+2}\}\Gamma_{n+2} \ge \eta_*\eta_{n+1}\Gamma_{n+2} \end{aligned} \tag{26}$$

where $\eta_* := \inf_{k \ge 2} \eta_{k+1}/\eta_k > 0$ by Assumption 15. Define the indices $\ell_n := 2n - 1$ for $n \ge 1$ and let

$$T_n := \eta_* \min\{\mu, Z_{\ell_n}\Gamma_{\ell_n}\}$$

for all $n \ge 2$. Define the σ-algebras $\mathscr{G}_n := \mathscr{F}_{\ell_n}$ for $n \ge 1$ and observe that $\mathbb{E}\left[T_{n+1} \;\middle|\; \mathscr{G}_n \right] \ge \eta_*\mu\delta/2$ by (25). Construct a martingale starting from $Y_1 \equiv 0$ and having the differences $\mathrm{d}Y_{n+1} := \eta_{\ell_n+1}(T_{n+1} - \mathbb{E}\left[T_{n+1} \;\middle|\; \mathscr{G}_n \right])$. The martingale Y_n converges to an a.s. finite limit Y_∞ as in Theorem 18.

Define also $\eta^* := \sup_{k \ge 2} \eta_{k+1}/\eta_k < \infty$ and $\kappa := \inf_{k \ge 2} 1 - \eta_k > 0$, and let

$$b := \frac{\kappa\eta_*\mu\delta}{8\eta^*} > 0.$$

[3] by possibly augmenting the probability space; see (Athreya and Ney 1978; Nummelin 1978).

Denote $S_n^{(w)} := \inf_{v\in\mathscr{S}(w)} v^T S_n v$ and define the stopping times $\tau_1 \equiv 1$ and for $k \geq 2$ through

$$\tau_k := \inf\{n > \tau_{k-1} : S_{\ell_n}^{(w)} \leq b,\, S_{\ell_{n-1}}^{(w)} > b\}$$

with the convention $\inf\emptyset = \infty$. That is, τ_k record the times when $S_{\ell_n}^{(w)}$ enters $(0, b]$. Using τ_k, define the latest such time up to n by $\sigma_n := \sup\{\tau_k : k \geq 1,\, \tau_k \leq n\}$.

Observe that for any $n \geq 2$ such that $S_{\ell_n}^{(w)} \leq b$, one may write

$$\begin{aligned}
S_{\ell_n}^{(w)} &= S_{\ell_{\sigma_n}}^{(w)} + \sum_{k=\sigma_n}^{n-1} \left(D_{\ell_k+2} - \eta_{\ell_k+1} S_{\ell_k}^{(w)} - \eta_{\ell_k+2} S_{\ell_k+1}^{(w)} \right) \\
&\geq S_{\ell_{\sigma_n}}^{(w)} + \sum_{k=\sigma_n}^{n-1} \left(\eta_{\ell_k+1} T_{k+1} - \eta_{\ell_k+1} b - \eta_{\ell_k+2} \kappa^{-1} b \right) \\
&\geq S_{\ell_{\sigma_n}}^{(w)} + \sum_{k=\sigma_n}^{n-1} \eta_{\ell_k+1} \left(T_{k+1} - \frac{\eta_* \mu \delta}{4} \right)
\end{aligned}$$

by (26) and since for all $k \in [\sigma_n, n-1]$ one may estimate $S_{\ell_k+1}^{(w)} \leq (1 - \eta_{\ell_k+1})^{-1} S_{\ell_{k+1}}^{(w)} \leq \kappa^{-1} b$.

That is, for any $n \geq 2$ such that $S_{\ell_n}^{(w)} \leq b$

$$\begin{aligned}
S_{\ell_n}^{(w)} &\geq S_{\ell_{\sigma_n}}^{(w)} + (Y_n - Y_{\sigma_n}) + \sum_{k=\sigma_n}^{n-1} \eta_{\ell_k+1} \left(\mathbb{E}\left[T_{k+1} \,\middle|\, \mathscr{G}_k \right] - \frac{\eta_* \delta \mu}{4} \right) \\
&\geq S_{\ell_{\sigma_n}}^{(w)} + (Y_n - Y_{\sigma_n}) + \frac{\eta_* \delta \mu}{4} \sum_{k=\sigma_n}^{n-1} \eta_{\ell_k+1}.
\end{aligned}$$

As in the proof of Theorem 19, this is sufficient to find a $\varepsilon > 0$ such that

$$\liminf_{n\to\infty} S_n^{(w)} \geq \varepsilon.$$

Finally, take a finite number of unit vectors $w_1, \ldots, w_N \in \mathscr{S}^d$ such that the corresponding segments $\mathscr{S}(w_1), \ldots, \mathscr{S}(w_N)$ cover $\mathscr{S}^d$. Then,

$$\liminf_{n\to\infty} \inf_{v\in\mathscr{S}^d} v^T S_n v = \liminf_{n\to\infty} \min\left\{ S_n^{(w_1)}, \ldots, S_n^{(w_N)} \right\} \geq \varepsilon. \qquad \square$$

Lemma 25. *Suppose $\mathscr{F}_n \subset \mathscr{F}_{n+1}$ are σ-algebras, and G_{n+1} and E_{n+1} are $\mathscr{F}_{n+1}$-measurable random variables, satisfying $\|E_{n+1}\| \leq M$ for some constant $M < \infty$. Moreover, U_{n+2} is a random variable independent of $\mathscr{F}_{n+1}$, having a distribution ν fulfilling the conditions in Assumption 22.*

Let $\mathscr{S}^d := \{u \in \mathbb{R}^d : \|u\| = 1\}$ stand for the unit sphere and denote by $\mathscr{S}(w,\gamma) := \{v \in \mathscr{S}^d : \|w - v\| \leq \gamma\}$ the segment of the unit sphere centred at $w \in \mathscr{S}^d$ and having the radius $\gamma > 0$. There exist constants $\gamma, \mu > 0$ such that

$$\mathbb{P}\left(\inf_{v\in\mathscr{S}(w,\gamma)} \left(|v^T G_{n+1}|^2 + |v^T (U_{n+2} + \lambda G_{n+1} + E_{n+1})|^2 \right) > \mu \,\middle|\, \mathscr{F}_n \right) \geq \frac{1}{2}.$$

for any $w \in \mathscr{S}^d$ and any constant $\lambda \in (0,1)$, almost surely.

Proof. Since ν is absolutely continuous with respect to the Lebesgue measure, one can show that there exist values $b, \gamma > 0$ such that

$$\inf_{w \in \mathscr{S}^d} \inf_{e \in B(0,M)} \nu \left(\{ u \in \mathbb{R}^d : \inf_{v \in \mathscr{S}(w,\gamma)} |v^T(u+e)| > b \} \right) \geq \frac{1}{2} \tag{27}$$

where $B(0,M) := \{y \in \mathbb{R}^d : \|y\| \leq M\}$ denotes a centred ball of radius M. Hereafter, fix $\gamma, b > 0$ such that (27) holds and let $a := b/2$.

Fix a unit vector $w \in \mathscr{S}^d$ and consider the set

$$\begin{aligned} A &:= \left\{ \inf_{v \in \mathscr{S}(w,\gamma)} \left(|v^T G_{n+1}|^2 + |v^T(U_{n+2} + \lambda G_{n+1} + E_{n+1})|^2 \right) \leq a^2 \right\} \\ &\subset \left\{ \inf_{v \in \mathscr{S}(w,\gamma) : |v^T G_{n+1}| \leq a} |v^T(U_{n+2} + \lambda G_{n+1} + E_{n+1})| \leq a \right\} \\ &\subset \left\{ \inf_{v \in \mathscr{S}(w,\gamma) : |v^T G_{n+1}| \leq a} |v^T(U_{n+2} + E_{n+1})| - \lambda |v^T G_{n+1}| \leq a \right\} \\ &\subset \left\{ \inf_{v \in \mathscr{S}(w,\gamma)} |v^T(U_{n+2} + E_{n+1})| \leq 2a \right\}. \end{aligned}$$

Since U_{n+2} is independent of $\mathscr{F}_{n+1}$, and since E_{n+1} is $\mathscr{F}_{n+1}$-measurable, one may estimate

$$\begin{aligned} \mathbb{P}\left(A^{\complement} \,\middle|\, \mathscr{F}_n \right) &\geq \mathbb{E}\left[\inf_{e \in B(0,M)} \mathbb{P}\left(\inf_{v \in \mathscr{S}(w,\gamma)} |v^T(U_{n+2} + e)| > 2a \,\middle|\, \mathscr{F}_{n+1} \right) \,\middle|\, \mathscr{F}_n \right] \\ &= \inf_{e \in B(0,M)} \nu \left(\{ u \in \mathbb{R}^d : \inf_{v \in \mathscr{S}(w,\gamma)} |v^T(u+e)| > b \} \right) \geq \frac{1}{2} \end{aligned}$$

by (27), almost surely, concluding the proof by $\mu := a^2$. □

Corollary 26. *Assume π is bounded, stays bounded away from zero on compact sets, is differentiable on the tails, and has regular contours, that is,*

$$\liminf_{\|x\| \to \infty} \frac{x}{\|x\|} \cdot \frac{\nabla \pi(x)}{\|\nabla \pi(x)\|} < 0. \tag{28}$$

Let $(X_n, M_n, S_n)_{n \geq 1}$ be an AM process as defined in Section 2 using a mixture proposal (5) *with a mixing weight satisfying $\beta \in (0,1)$ and the density q_{fix} is bounded away from zero in some neighbourhood of the origin. Moreover, suppose that the adaptation weights $(\eta_n)_{n \geq 2}$ satisfy Assumptions 15 and 17. Then,*

$$\liminf_{n \to \infty} \inf_{w \in \mathscr{S}^d} w^T S_n w > 0.$$

Proof. In light of Theorem 24, it is sufficient to check Assumption 22, or in fact the conditions in Remark 23. Let $L > 0$ be sufficiently large so that $\inf_{\|x\| \geq L} \frac{x}{\|x\|} \cdot \frac{\nabla \pi(x)}{\|\nabla \pi(x)\|} < 0$. Jarner and Hansen (Jarner and Hansen 2000, proof of Theorem 4.3) show that there is an $\epsilon' > 0$ and $K > 0$ such that the cone

$$E(x) := \left\{ x - au : 0 < a < K,\, u \in \mathscr{S}^d,\, \left\| u - \frac{x}{\|x\|} \right\| \leq \epsilon' \right\}$$

is contained in the set $A(x) := \{y \in \mathbb{R}^d : \pi(y) \geq \pi(x)\}$, for all $\|x\| \geq L$.

Let $r' > 0$ be sufficiently small to ensure that $\inf_{\|z\|\leq r'} q_{\text{fix}}(z) \geq \delta' > 0$. There is a $r = r(\epsilon', K) \in (0, r'/2)$ and measurable $\xi : \mathbb{R}^d \to \mathbb{R}^d$ such that $\|\xi(x) - x\| \leq r'/2$ and the ball $B(x,r) := \{y : \|y - \xi(x)\| \leq r\}$ is contained in the cone $E(x)$. Define $\nu(x) := c_r^{-1} \mathbb{1}_{B(0,r)}(x)$ where $c_r := |B(0,r)|$ is the Lebesgue measure of $B(0,r)$, and let $\xi(x) := x$ for the remaining $\|x\| < L$. Now, we have for $\|x\| \geq L$ that

$$\beta q_{\text{fix}}(x-y) \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} \geq \beta \delta' c_r \nu(y - \xi).$$

Since π is bounded and bounded away from zero on compact sets, the ratio $\pi(y)/\pi(x) \geq \delta'' > 0$ for all $x, y \in B(0, L + r')$ with $\|x - y\| \leq r'$. Therefore, for all $\|x\| < L$, it holds that

$$\beta q_{\text{fix}}(x-y) \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} \geq \beta \delta' \delta'' c_r \nu(y - x). \qquad \square$$

Remark 27. The conditions of Corollary 26 are fulfilled by many practical densities π (see Jarner and Hansen (2000) for examples), and are fairly easy to verify in practice. Assumption 22 holds, however, more generally, excluding only densities with unbounded density or having irregular contours.

Remark 28. It is not necessary for Theorem 24 and Corollary 26 to hold that the adaptive proposal densities $\{\tilde{q}_s\}_{s \in \mathscr{C}^d}$ have the specific form discussed in Section 2. The results require only that a suitable fixed proposal component is used so that Assumption 22 holds. In Theorem 29 below, however, the structure of $\{\tilde{q}_s\}_{s \in \mathscr{C}^d}$ is required.

Let us record the following ergodicity result, which is a counterpart to (Saksman and Vihola 2010, Theorem 10) formulating a a strong law of large numbers for the original algorithm (S1)–(S3) with the covariance parameter (1).

Theorem 29. *Suppose the target density π is continuous and differentiable, stays bounded away from zero on compact sets and has super-exponentially decaying tails with regular contours,*

$$\limsup_{\|x\|\to\infty} \frac{x}{\|x\|^\rho} \cdot \nabla \log \pi(x) = -\infty \qquad \text{and} \qquad \limsup_{\|x\|\to\infty} \frac{x}{\|x\|} \cdot \frac{\nabla \pi(x)}{\|\nabla \pi(x)\|} < 0,$$

respectively, for some $\rho > 1$.

Let $(X_n, M_n, S_n)_{n\geq 1}$ be an AM process as defined in Section 2 using a mixture proposal $q_s(z) = (1-\beta)\tilde{q}_s(z) + \beta q_{fix}(z)$ where $\tilde{q}_s$ stands for a zero-mean Gaussian density with covariance s, the mixing weight satisfies $\beta \in (0,1)$ and the density q_{fix} is bounded away from zero in some neighbourhood of the origin. Moreover, suppose that the adaptation weights $(\eta_n)_{n\geq 2}$ satisfy Assumption 17.

Then, for any function $f : \mathbb{R}^d \to \mathbb{R}$ with $\sup_{x\in\mathbb{R}^d} \pi^\gamma(x)|f(x)| < \infty$ for some $\gamma \in (0, 1/2)$,

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

almost surely.

Proof. The conditions of Corollary 26 are satisfied, implying that for any $\epsilon > 0$ there is a $\kappa = \kappa(\epsilon) > 0$ such that $\mathbb{P}\big(\inf_n \lambda_{\min}(S_n) \geq \kappa\big) \geq 1 - \epsilon$ where $\lambda_{\min}(s)$ denotes the smallest eigenvalue of s. By (Saksman and Vihola 2010, Proposition 15), there is a compact set $C_\kappa \subset \mathbb{R}^d$, a probability measure ν_κ on C_κ, and $b_\kappa < \infty$ such that for all $s \in \mathscr{C}^d$ with $\lambda_{\min}(s) \geq \kappa$, it holds that

$$P_{\tilde{q}_s} V(x) \leq \lambda_s V(x) + b \mathbb{1}_{C_\kappa}(x), \qquad \forall x \in \mathbb{R}^d \tag{29}$$

$$P_{\tilde{q}_s}(x, A) \geq \delta_s \nu(A) \qquad \forall x \in C_\kappa \tag{30}$$

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

$$(1 - \lambda_s)^{-1} \vee \delta_s^{-1} \leq c_1 \det(s)^{1/2} \tag{31}$$

for some constant $c_1 \geq 1$. Likewise, there is a compact $D_f \subset \mathbb{R}^d$, a probability measure μ_f on D_f, and constants $b_f < \infty$ and $\lambda_f, \delta_f \in (0,1)$, so that (29) and (30) hold with P_f (Jarner and Hansen 2000, Theorem 4.3). Put together, (29) and (30) hold for P_{q_s} for all $s \in \mathscr{C}^d$ with $\lambda_{\min}(s) \geq \kappa$, perhaps with different constants, but satisfying a bound (31), with another $c_2 \geq c_1$.

The rest of the proof follows as in Theorem 21 by construction of an auxiliary process $(\tilde{X}_n, \tilde{M}_n, \tilde{S}_n)_{n \geq 1}$ truncated so that for given $\varepsilon > 0$, $\kappa \leq \lambda_{\min}(\tilde{S}_n) \leq a n^{\varepsilon}$ and $|\tilde{M}_n| \leq a n^{\varepsilon}$ and where the constant $a = a(\varepsilon, \kappa)$ is chosen so that the truncated process coincides with the original AM process with probability $\geq 1 - 2\epsilon$. Theorem 2 of Saksman and Vihola (2010) ensures that the strong law of large numbers holds for the constrained process, and letting $\epsilon \to 0$ implies the claim. □

Remark 30. In the case $\eta_n := n^{-1}$, Theorem 29 implies that with probability one, $M_n \to m_\pi := \int x \pi(x) \mathrm{d}x$ and $S_n \to s_\pi := \int x x^T \pi(x) \mathrm{d}x - m_\pi m_\pi^T$, the true mean and covariance of π, respectively.

Remark 31. Theorem 29 holds also when using multivariate Student distributions $\{\tilde{q}_s\}_{s \in \mathscr{C}^d}$, as (Vihola 2009, Proposition 26 and Lemma 28) extend the result in Saksman and Vihola (2010) to cover this case.

Acknowledgements

The author thanks Professor Eero Saksman for discussions and helpful comments on the manuscript.

References

C. Andrieu and É. Moulines. On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.*, 16(3):1462–1505, 2006. MR2260070

C. Andrieu and C. P. Robert. Controlled MCMC for optimal sampling. Technical Report Ceremade 0125, Université Paris Dauphine, 2001.

C. Andrieu and J. Thoms. A tutorial on adaptive MCMC. *Statist. Comput.*, 18(4):343–373, Dec. 2008. MR2461882

Y. Atchadé and G. Fort. Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli*, 16(1):116–154, Feb. 2010. MR2648752

Y. F. Atchadé and J. S. Rosenthal. On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, 11 (5):815–828, 2005. MR2172842

K. B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.*, 245:493–501, 1978. MR0511425

Y. Bai, G. O. Roberts, and J. S. Rosenthal. On the containment condition for adaptive Markov chain Monte Carlo algorithms. Preprint, July 2008. URL `http://probability.ca/jeff/research.html`. MR2418242

C. G. Esseen. On the Kolmogorov-Rogozin inequality for the concentration function. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 5(3):210–216, Sept. 1966. MR0205297

H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001. MR1828504

P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980. ISBN 0-12-319350-8. MR0624435

S. F. Jarner and E. Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.*, 85:341–361, 2000. MR1731030

E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 43(3):309–318, Dec. 1978. MR0501353

G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004. MR2095565

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. MR2340211

G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *J. Comput. Graph. Statist.*, 18(2): 349–367, 2009.

B. A. Rogozin. An estimate for concentration functions. *Theory Probab. Appl.*, 6(1):94–97, Jan. 1961. MR0131893

E. Saksman and M. Vihola. On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab.*, 20(6):2178–2203, Dec. 2010.

M. Vihola. On the stability and ergodicity of an adaptive scaling Metropolis algorithm. Preprint, arXiv:0903.4061v2, Mar. 2009.

A Lemmas for Section 3.2

Let us start by establishing some properties for a weight sequence $(\eta_n)_{n\geq 1}$ satisfying Assumption 11.

Lemma 32. *Suppose $(\eta_n)_{n\geq 1}$ satisfies Assumption 11. Then,* *(a)* $(\eta_{n+1}/\eta_n)_{n\geq m'}$ *is increasing with* $\eta_{n+1}/\eta_n \to 1$ *and*

(b) $\eta_{n+1}^{-1/2} - \eta_n^{-1/2} \to 0$.

Proof. Define $a_n := \eta_n^{-1/2}$ for all $n \geq m'$. By Assumption 11 (i) $(a_n)_{n\geq m'}$ is increasing and by Assumption 11 (ii), $(\Delta a_n)_{n\geq m'+1}$ is decreasing, where $\Delta a_n := a_n - a_{n-1}$. One can write

$$\frac{a_n}{a_{n+1}} = \frac{1}{1+\frac{\Delta a_{n+1}}{a_n}} \geq \frac{1}{1+\frac{\Delta a_n}{a_{n-1}}} = \frac{a_{n-1}}{a_n}$$

implying that $(\eta_{n+1}/\eta_n)_{n\geq m'}$ is increasing. Denote $c = \lim_{n\to\infty} \eta_{n+1}/\eta_n \leq 1$. It holds that $\eta_{m'+k} \leq c\eta_{m'+k-1} \leq \cdots \leq c^k \eta_{m'}$. If $c < 1$, then $\sum_n \eta_n < \infty$ contradicting Assumption 11 (iii), so c must be one, establishing (a).

From (a), one obtains

$$\frac{\eta_{n+1}^{-1/2} - \eta_n^{-1/2}}{\eta_n^{-1/2}} = \left(\frac{\eta_n}{\eta_{n+1}}\right)^{1/2} - 1 \to 0$$

implying (b). □

Lemma 33. *Suppose* $m_1 \geq 1$, $g_{m_1} \geq 0$, *the sequence* $(\eta_n)_{n\geq m_1}$ *satisfies Assumption 11 and* $\tilde{\theta} > 0$ *is a constant. The sequence* $(g_n)_{n>m_1}$ *defined through*

$$g_{n+1} := \eta_{n+1}^{1/2}\left(\frac{(1-\eta_n)^3}{\eta_n}\frac{g_n}{g_n+\eta_n^{-1/2}} + \tilde{\theta}^2\right)$$

satisfies $\lim_{n\to\infty} g_n = \tilde{\theta}$.

Proof. Define the functions $f_n : \mathbb{R}_+ \to \mathbb{R}_+$ for $n \geq m_1 + 1$ by

$$f_{n+1}(x) := \eta_{n+1}^{1/2}\left(\frac{(1-\eta_n)^3}{\eta_n}\frac{x}{x+\eta_n^{-1/2}} + \tilde{\theta}^2\right).$$

The functions f_n are contractions on $[0,\infty)$ with contraction coefficient $q_n := (1-\eta_n)^3$ since for all $x, y \geq 0$

$$\begin{aligned}
\left|f_{n+1}(x) - f_{n+1}(y)\right| &= \eta_{n+1}^{1/2}\frac{(1-\eta_n)^3}{\eta_n}\left|\frac{x}{x+\eta_n^{-1/2}} - \frac{y}{y+\eta_n^{-1/2}}\right| \\
&= \left(\frac{\eta_{n+1}}{\eta_n}\right)^{1/2}\frac{(1-\eta_n)^3}{\eta_n}\left|\frac{x-y}{(x+\eta_n^{-1/2})(y+\eta_n^{-1/2})}\right| \\
&\leq \left(\frac{\eta_{n+1}}{\eta_n}\right)^{1/2}(1-\eta_n)^3\left|x-y\right| \leq q_{n+1}\left|x-y\right|
\end{aligned}$$

where the second inequality holds since $\eta_{n+1} \leq \eta_n$.

The fixed point of f_{n+1} can be written as

$$x_{n+1}^* := \frac{1}{2}\left(-\xi_{n+1} + \sqrt{\xi_{n+1}^2 + \mu_{n+1}}\right)$$

where

$$\xi_{n+1} := \eta_n^{-1/2} - \eta_{n+1}^{1/2}\eta_n^{-1}(1-\eta_n)^3 - \eta_{n+1}^{1/2}\tilde{\theta}^2$$
$$\mu_{n+1} := 4\eta_n^{-1/2}\eta_{n+1}^{1/2}\tilde{\theta}^2.$$

Lemma 32 (a) implies $\mu_{n+1} \to 4\tilde{\theta}^2$. Moreover,

$$\begin{aligned}\xi_{n+1} &= \eta_n^{-1/2} - \eta_{n+1}^{1/2}\eta_n^{-1} + \eta_{n+1}^{1/2}(3 - 3\eta_n + \eta_n^2 - \tilde{\theta}^2) \\ &= \left(\frac{\eta_{n+1}}{\eta_n}\right)^{1/2}\left(\eta_{n+1}^{-1/2} - \eta_n^{-1/2}\right) + \eta_{n+1}^{1/2}(3 - 3\eta_n + \eta_n^2 - \tilde{\theta}^2).\end{aligned}$$

Therefore, by Assumption 11 (i) and Lemma 32, $\xi_{n+1} \to 0$ and consequently the fixed points satisfy $x_n^* \to \tilde{\theta}$.

Consider next the consecutive differences of the fixed points. Using the mean value theorem and the triangle inequality, write

$$\begin{aligned}2\left|x_{n+1}^* - x_n^*\right| &\le \left|\xi_{n+1} - \xi_n\right| + \frac{1}{2\sqrt{\tau_n}}\left|\xi_{n+1}^2 - \xi_n^2 + \mu_{n+1} - \mu_n\right| \\ &\le \left|\xi_{n+1} - \xi_n\right| + \frac{\tau_n'}{\sqrt{\tau_n}}\left|\xi_{n+1} - \xi_n\right| + \frac{1}{2\sqrt{\tau_n}}\left|\mu_{n+1} - \mu_n\right| \\ &\le c_1\left|\xi_{n+1} - \xi_n\right| + c_1\left|\mu_{n+1} - \mu_n\right|\end{aligned}$$

where the value of τ_n is between $\xi_{n+1}^2 + \mu_{n+1}$ and $\xi_n^2 + \mu_n$ converging to $4\tilde{\theta}^2 > 0$, the value of τ_n' is between $|\xi_{n+1}|$ and $|\xi_n|$ converging to zero, and $c_1 > 0$ is a constant.

The differences of the latter terms satisfy for all $m \ge m'$

$$\begin{aligned}\sum_{n=m'}^{m}\left|\mu_{n+1} - \mu_n\right| &= 4\tilde{\theta}^2\sum_{n=m'}^{m}\left[\left(\frac{\eta_{n+1}}{\eta_n}\right)^{1/2} - \left(\frac{\eta_n}{\eta_{n-1}}\right)^{1/2}\right] \\ &\le 4\tilde{\theta}^2\left[1 - \left(\frac{\eta_{m'}}{\eta_{m'-1}}\right)^{1/2}\right] \le 4\tilde{\theta}^2.\end{aligned}$$

by Assumption 11 (ii) and Lemma 32 (a). For the first term, let us estimate

$$\begin{aligned}\left|\xi_{n+1} - \xi_n\right| \le &\left|\left(\frac{\eta_{n+1}}{\eta_n}\right)^{1/2}\left(\eta_{n+1}^{-1/2} - \eta_n^{-1/2}\right) - \left(\frac{\eta_n}{\eta_{n-1}}\right)^{1/2}\left(\eta_n^{-1/2} - \eta_{n-1}^{-1/2}\right)\right| \\ &+ \left|3 - \tilde{\theta}^2\right|\left|\eta_n^{1/2} - \eta_{n+1}^{1/2}\right| + \left|\eta_{n+1}^{1/2}(3\eta_n - \eta_n^2) - \eta_n^{1/2}(3\eta_{n-1} - \eta_{n-1}^2)\right|.\end{aligned}$$

Assumption 11 (i) implies that $\eta_n^{1/2} - \eta_{n+1}^{1/2} \ge 0$ for $n \ge m'$ and hence $\sum_{n=m'}^{m}\left|\eta_n^{1/2} - \eta_{n+1}^{1/2}\right| \le \eta_{m'}^{1/2}$ for any $m \ge m'$. Since the function $(x, y) \mapsto x(3y - y^2)$ is Lipschitz on $[0, 1]^2$, there is a constant c_2 independent of n such that $\left|\eta_{n+1}^{1/2}(3\eta_n - \eta_n^2) - \eta_n^{1/2}(3\eta_{n-1} - \eta_{n-1}^2)\right| \le c_2\left(|\eta_{n+1}^{1/2} - \eta_n^{1/2}| + |\eta_n - \eta_{n-1}|\right)$, and a similar argument shows that

$$\sum_{n=m'}^{m}\left|\eta_{n+1}^{1/2}(3\eta_n - \eta_n^2) - \eta_n^{1/2}(3\eta_{n-1} - \eta_{n-1}^2)\right| \le c_3 < \infty.$$

One can also estimate

$$\left|\left(\frac{\eta_{n+1}}{\eta_n}\right)^{1/2}\left(\eta_{n+1}^{-1/2}-\eta_n^{-1/2}\right)-\left(\frac{\eta_n}{\eta_{n-1}}\right)^{1/2}\left(\eta_n^{-1/2}-\eta_{n-1}^{-1/2}\right)\right|$$
$$\le c_4\left|\left(\frac{\eta_{n+1}}{\eta_n}\right)^{1/2}-\left(\frac{\eta_n}{\eta_{n-1}}\right)^{1/2}\right|+c_4\left|\left(\eta_{n+1}^{-1/2}-\eta_n^{-1/2}\right)-\left(\eta_n^{-1/2}-\eta_{n-1}^{-1/2}\right)\right|$$

yielding by Assumption 11 (ii) and Lemma 32 that $\sum_{n=m'}^{m}|\xi_{n+1}-\xi_n|\le c_5$ for all $m\ge m'$, with a constant $c_5<\infty$. Combining the above estimates, the fixed point differences satisfy

$$\sum_{n=m'}^{m}|x_{n+1}^*-x_n^*|<\infty.$$

Fix a $\delta>0$ and let $n_\delta>m_1$ be sufficiently large so that $\sum_{k=n_\delta+1}^{\infty}|x_{n+1}^*-x_n^*|\le\delta$ implying also that $|x_n^*-\tilde{\theta}|\le\delta$ for all $n\ge n_\delta$. Then, for $n\ge n_\delta$ one may write

$$\begin{aligned}\left|g_n-\tilde{\theta}\right| &\le \left|g_n-x_n^*\right|+\left|x_n^*-\tilde{\theta}\right|\le\left|f_n(g_{n-1})-f_n(x_n^*)\right|+\delta\\ &\le q_n\left|g_{n-1}-x_n^*\right|+\delta\le q_n\left|g_{n-1}-x_{n-1}^*\right|+\left|x_{n-1}^*-x_n^*\right|+\delta\\ &\le q_nq_{n-1}\left|g_{n-2}-x_{n-2}^*\right|+\left|x_{n-2}^*-x_{n-1}^*\right|+\left|x_{n-1}^*-x_n^*\right|+\delta\\ &\le\cdots\le\left(\prod_{k=n_\delta+1}^{n}q_k\right)\left|g_{n_\delta}-x_{n_\delta}^*\right|+2\delta.\end{aligned}$$

Since $\log\prod_{k=n_\delta+1}^{n}q_k=3\sum_{k=n_\delta+1}^{n}\log(1-\eta_{k-1})\le-3\sum_{k=n_\delta}^{n-1}\eta_k\to-\infty$ as $n\to\infty$ by Assumption 11 (iii), it holds that $(\prod_{k=n_\delta+1}^{n}q_k)|g_{n_\delta}-x_{n_\delta}^*|\to0$. That is, $|g_n-\tilde{\theta}|\le3\delta$ for any sufficiently large n, and since $\delta>0$ was arbitrary, $g_n\to\tilde{\theta}$. □

B Lemmas for Section 3.3

Proof of Lemma 13. Equation (14) follows directly by writing

$$\begin{aligned}u^TS_{n+1}u&=(1-\eta_{n+1})u^TS_nu+\eta_{n+1}u^T(X_{n+1}-M_n)(X_{n+1}-M_n)^Tu\\&=[1+\eta_{n+1}(Z_{n+1}^2-1)]u^TS_nu.\end{aligned}$$

For $n\ge2$, write using the above equation

$$\begin{aligned}Z_{n+1}&=\theta u^T\frac{S_n^{1/2}W_{n+1}}{\|S_n^{1/2}u\|}+(1-\eta_n)u^T\frac{X_n-M_{n-1}}{\|S_n^{1/2}u\|}\\&=\theta\frac{u^TS_n^{1/2}}{\|S_n^{1/2}u\|}W_{n+1}+(1-\eta_n)\left(\frac{u^TS_{n-1}u}{u^TS_nu}\right)^{1/2}Z_n\\&=\theta\tilde{W}_{n+1}+(1-\eta_n)\left(\frac{1}{1+\eta_n(Z_n^2-1)}\right)^{1/2}Z_n\end{aligned}$$

where $(\tilde{W}_{n+1}:=\|S_n^{1/2}u\|^{-1}u^TS_n^{1/2}W_{n+1}$ are non-degenerate i.i.d. random variables by Lemma 34 below. This establishes (15). □

Lemma 34. *Assume $u \in \mathbb{R}^d$ is a non-zero vector, $(W_n)_{n\geq 1}$ are independent random variables following a common spherically symmetric non-degenerate distribution in $\mathbb{R}^d$. Assume also that S_n are symmetric and positive definite random matrices taking values in $\mathbb{R}^{d\times d}$ measurable with respect $\mathscr{F}_n := \sigma(W_1, \ldots, W_n)$. Then the random variables $(\tilde{W}_n)_{n\geq 2}$ defined through*

$$\tilde{W}_{n+1} := \frac{u^T S_n^{1/2}}{\|S_n^{1/2} u\|} W_{n+1}$$

are i.i.d. non-degenerate real-valued random variables.

Proof. Choose a measurable $A \subset \mathbb{R}$, denote $T_n := \|S_n^{1/2} u\|^{-1} S_n^{1/2} u$ and define $A_n := \{x \in \mathbb{R}^d : T_n^T x \in A\}$. Let R_n be a rotation matrix such that $R_n^T T_n = e_1 := (1, 0, \ldots, 0) \in \mathbb{R}^d$. Since W_{n+1} is independent of $\mathscr{F}_n$, we have

$$\begin{aligned} \mathbb{P}(\tilde{W}_{n+1} \in A \mid \mathscr{F}_n) &= \mathbb{P}(W_{n+1} \in A_n \mid \mathscr{F}_n) = \mathbb{P}(R_n W_{n+1} \in A_n \mid \mathscr{F}_n) \\ &= \mathbb{P}(e_1^T W_{n+1} \in A \mid \mathscr{F}_n) = \mathbb{P}(e_1^T W_1 \in A) \end{aligned}$$

by the rotational invariance of the distribution of $(W_n)_{n\geq 1}$. Since the common distribution of $(W_n)_{n\geq 1}$ is non-degenerate, so is the distribution of $e_1^T W_1$. □

Remark 35. Notice particularly that if $(W_n)_{n\geq 2}$ in Lemma 34 are standard Gaussian vectors in $\mathbb{R}^d$ then $(\tilde{W}_n)_{n\geq 2}$ are standard Gaussian random variables.

C The Kolmogorov-Rogozin inequality

Define the concentration function $Q(X; \lambda)$ of a random variable X by

$$Q(X; \lambda) := \sup_{x\in\mathbb{R}} \mathbb{P}(X \in [x, x + \lambda])$$

for all $\lambda \geq 0$.

Theorem 36. *Let $X_1, X_2, \ldots$ be mutually independent random variables. There is a universal constant $c > 0$ such that*

$$Q\left(\sum_{k=1}^{n} X_k; L\right) \leq \frac{cL}{\lambda}\left(\sum_{k=1}^{n} \big(1 - Q(X_k; \lambda)\big)\right)^{-1/2}$$

for all $L \geq \lambda > 0$.

Proof. Rogozin's original work Rogozin (1961) uses combinatorial results, and Esseen's alternative proof Esseen (1966) is based on characteristic functions. □

D A coupling construction

Theorem 37. *Suppose μ and ν are probability measures and the random variable $X \sim \mu$. Then, possibly by augmenting the probability space, there is another random variable Y such that $Y \sim \nu$ and $\mathbb{P}(X = Y) = 1 - \|\mu - \nu\|$.*

Proof (adopted from Theorem 3 in (Roberts and Rosenthal 2004)). Define the measure $\rho := \mu + \nu$, and the densities $g := \mathrm{d}\mu/\mathrm{d}\rho$ and $h := \mathrm{d}\nu/\mathrm{d}\rho$, existing by the Radon-Nikodym theorem. Let us introduce two auxiliary variables U and Z independent of each other and X, whose existence is ensured by possible augmentation of the probability space. Then, Y is defined through

$$Y = \mathbb{1}_{\{U \le r(X)\}} X + \mathbb{1}_{\{U > r(X)\}} Z$$

where the 'coupling probability' r is defined as $r(y) := \min\{1, h(y)/g(y)\}$ whenever $g(y) > 0$ and $r(y) := 1$ otherwise. The variable U is uniformly distributed on $[0,1]$. If $r(y) = 1$ for ρ-almost every y, then the choice of Z is irrelevant, $\mu = \nu$, and the claim is trivial. Otherwise, the variable Z is distributed following the 'residual measure' ξ given as

$$\xi(A) := \frac{\int_A \max\{0, h - g\}\mathrm{d}\rho}{\int \max\{0, h - g\}\mathrm{d}\rho}.$$

Observe that $\int \max\{0, h - g\}\mathrm{d}\rho = \int \max\{0, g - h\}\mathrm{d}\rho > 0$ in this case, so ξ is a well defined probability measure.

Let us check that $Y \sim \nu$,

$$\begin{aligned}
\mathbb{P}(Y \in A) &= \int_A r\mathrm{d}\mu + \xi(A) \int (1 - r)\mathrm{d}\mu \\
&= \int_A \min\{g, h\}\mathrm{d}\rho + \xi(A) \int_{h<g} (g - h)\mathrm{d}\rho \\
&= \int_A \min\{g, h\} + \max\{0, h - g\}\rho(\mathrm{d}x) = \nu(A).
\end{aligned}$$

Moreover, by observing that $r(y) = 1$ in the support of ξ, one has

$$\mathbb{P}(X = Y) = \int r\mathrm{d}\mu = \int \min\{g, h\}\mathrm{d}\rho = 1 - \int_{g<h} (h - g)\mathrm{d}\rho = 1 - \|\nu - \mu\|$$

since $\int_{g<h}(h - g)\mathrm{d}\rho = \int_{h<g}(g - h)\mathrm{d}\rho = \sup_f \left|\int f(h - g)\mathrm{d}\rho\right| = \|\mu - \nu\|$ where the supremum taken over all measurable functions f taking values in $[0,1]$. □

E Proof of Lemma 20

Observe that without loss of generality it is sufficient to check the case $m = 0$ and $b = 1$, that is, consider the standard Laplace distribution $\pi(x) := \frac{1}{2}e^{-|x|}$.

Let $x > 0$ and start by writing

$$1 - \frac{P_s V(x)}{V(x)} = \int_{-x}^{x} a(x, y)\tilde{q}_s(y - x)\mathrm{d}y - \int_{|y|>x} b(x, y)\tilde{q}_s(y - x)\mathrm{d}y \tag{32}$$

where

$$a(x,y) := \left(1 - \sqrt{\frac{\pi(x)}{\pi(y)}}\right) = 1 - e^{-\frac{x-|y|}{2}} \qquad \text{and}$$

$$b(x,y) := \sqrt{\frac{\pi(y)}{\pi(x)}}\left(1 - \sqrt{\frac{\pi(y)}{\pi(x)}}\right) = e^{-\frac{|y|-x}{2}}\left(1 - e^{-\frac{|y|-x}{2}}\right).$$

Compute then that

$$\int_0^x a(x,y)\tilde{q}_s(y-x)\mathrm{d}y - \int_x^{2x} b(x,y)\tilde{q}_s(y-x)\mathrm{d}y = \int_0^x (1-e^{-\frac{z}{2}})^2\tilde{q}_s(z)\mathrm{d}z.$$

The estimates

$$\int_{-x}^0 a(x,y)\tilde{q}_s(y-x)\mathrm{d}y \geq \tilde{q}_s(2x)\int_0^x a(x,y)\mathrm{d}y = \tilde{q}_s(2x)\int_0^x (1-e^{-\frac{z}{2}})\mathrm{d}z$$
$$\int_{-\infty}^{-x} b(x,y)\tilde{q}_s(y-x)\mathrm{d}y \leq \tilde{q}_s(2x)\int_x^\infty b(x,y)\mathrm{d}y = \tilde{q}_s(2x)\int_0^\infty e^{-\frac{z}{2}}(1-e^{\frac{z}{2}})\mathrm{d}z$$

due to the non-increasing property of $\tilde{q}_s$ yield

$$\int_{-x}^0 a(x,y)\tilde{q}_s(y-x)\mathrm{d}y - \int_{-\infty}^{-x} b(x,y)\tilde{q}_s(y-x)\mathrm{d}y$$
$$\geq \tilde{q}_s(2x)\left[\int_0^x (1-e^{-\frac{z}{2}})^2\mathrm{d}z - \int_x^\infty e^{-\frac{z}{2}}\mathrm{d}z\right] > 0$$

for any sufficiently large $x > 0$. Similarly, one obtains

$$\frac{1}{2}\int_0^x (1-e^{-\frac{z}{2}})^2\tilde{q}_s(z)\mathrm{d}z - \int_{2x}^\infty b(x,y)q_s(y-x)\mathrm{d}y > 0$$

for large enough $x > 0$.

Summing up, letting $M > 0$ be sufficiently large, then for $x \geq M$ and $s \geq L > 0$

$$1 - \frac{P_sV(x)}{V(x)} \geq \frac{1}{2}\int_0^x (1-e^{-\frac{z}{2}})^2\tilde{q}_s(z)\mathrm{d}z \geq \frac{1}{2}\tilde{q}_s(M)\int_0^M (1-e^{-\frac{z}{2}})^2\mathrm{d}z$$
$$\geq c_1 s^{-1/2}\tilde{q}(\theta^{-1/2}s^{-1/2}M) \geq c_2 s^{-1/2}$$

for some constants $c_1, c_2 > 0$. The same inequality holds also for $-x \leq -M$ due to symmetry. The simple bound $P_sV(x) \leq 2V(x)$ observed from (32) with the above estimate establishes (22). The minorisation inequality (23) holds since for all $x \in C$ one may write

$$P_s(x,A) \geq \int_{A\cap C} \max\left\{1, \frac{\pi(y)}{\pi(x)}\right\}\tilde{q}_s(y-x)\mathrm{d}y$$
$$\geq \frac{\inf_{z\in C}\pi(z)}{\sup_z \pi(z)} \inf_{s\geq L, z, y\in C}\tilde{q}_s(z-y)\int_{A\cap C}\mathrm{d}y \geq c_3 s^{-1/2}\nu(A).$$

where $\nu(A) := |A\cap C|/|C|$ with $|\cdot|$ denoting the Lebesgue measure. □