Weak Convergence of Metropolis Algorithms for Non-*iid* Target Distributions

Mylène Bédard *

April 3, 2006

Abstract

In this paper, we shall optimize the efficiency of random walk Metropolis algorithms for multidimensional target distributions with scaling terms possibly depending on the dimension. We propose a method to determine the appropriate form for the scaling of the proposal distribution as a function of the dimension, which leads to the proof of an asymptotic diffusion theorem. We show that when there does not exist any component having a scaling term significantly smaller than the others, the asymptotically optimal acceptance rate is the well-known 0.234.

1 Introduction

The characteristic of Metropolis-Hastings algorithms ([12], [11]) resides in the necessity of choosing a proposal density for their implementation. When this proposal distribution is chosen such that the kernel driving the chain is a random walk, the algorithm is referred to as a random walk Metropolis (RWM) algorithm, which is the most commonly used class of Metropolis-Hastings algorithms. Their ease of implementation and wide applicability have conferred their popularity to RWM algorithms and they are frequently used nowadays by all levels of practitioners in various fields of application. Their versatility however implies that they are not problem-specific and their convergence can sometimes be lengthy, which calls for an optimization of their performance. Because the efficiency of Metropolis-Hastings algorithms depends crucially on the scaling of the proposal distribution, it is thus fundamental to judiciously choose this parameter.

*Department of Statistics, University of Toronto, Toronto, Ontario, Canada, M5S 3G3. Email: mylene@utstat.utoronto.ca

Informal guidelines for the optimal scaling problem have been proposed among others by [3] and [4], but the first theoretical results have been obtained by [14]. In particular, the authors considered d-dimensional target distributions with *iid* components and studied the asymptotic behavior (as $d \to \infty$) of RWM algorithms with Gaussian proposals. It was proved that under some regularity conditions for the target distribution, the asymptotic acceptance rate should be tuned to be approximately 0.234 for optimal performance of the algorithm. It was also shown that the correct proposal scaling is of the form ℓ^2/d for some constant ℓ as $d \to \infty$. The simplicity of the obtained asymptotically optimal acceptance rate makes these theoretical results extremely useful in practice. Optimal scaling issues have been explored by other authors, namely [15], [7], [6], [8] and [13]. A good review of general optimal scaling results is given in [16].

In this paper, we carry out a similar study for d-dimensional target distributions with independent components. The particularity of our model resides in the fact that the scaling term of each component is allowed to depend on the dimension of the target distribution. This results in a high instability of the scaling terms, as they are allowed to converge both to 0 and ∞ as the dimension increases. Despite the independence of the various target components, the disparities exhibited by the scaling terms constitute a critical distinction with the *iid* case. Furthermore, because Gaussian distributions are invariant to orthogonal transformations, the model studied also includes multivariate normal target distributions with correlated components.

We provide a necessary and sufficient condition under which the algorithm will admit the same limiting diffusion process and the same asymptotically optimal acceptance rate as those found in [14]. To this end, an appropriate rescaling of space and time allows us to obtain a nontrivial limiting process as $d \to \infty$. This is achieved in the first place by presenting a method to determine the appropriate scaling form of the proposal distribution as $d \to \infty$, which is now different from the *iid* case. Then, by verifying $\mathcal{L}^1$ convergence of generators, we prove that the sequence of stochastic processes formed by say the i-th component of each Markov chain converges to a Langevin diffusion process with a certain speed measure. Obtaining the asymptotically optimal acceptance rate is thus a simple matter of optimizing the speed measure of the diffusion.

The paper is structured as follows. In Section 2, we outline the MCMC setup and introduce some definitions. Section 3 aims to describe the target distribution setting, while Section 4 is used to define the proposal distribution and its optimal scaling form. The main results are presented in Section 5. Inhomogeneous targets are then discussed in Section 6, along with some extensions. We prove the theorems in Section 7 using lemmas proved in Sections 8 and 9, finally concluding the paper with a discussion in the last section.

2 Algorithm and Definitions

Metropolis-Hastings algorithms provide a way to generate a Markov chain $\mathbf{X}_0, \mathbf{X}_1, \ldots$ having the target distribution as a stationary distribution. In particular, suppose that π is a d-dimensional probability density of interest with respect to some measure μ. Also, let $Q(\mathbf{x}, \cdot)$ be some chosen proposal kernel having density $q(\mathbf{x}, \mathbf{y})$ with respect to the same measure μ. The Metropolis-Hastings algorithm thus proceeds as follows. Given $\mathbf{X}_t$, the state of the chain at time t, a value $\mathbf{Y}_{t+1}$ is generated from $q(\mathbf{X}_t, \mathbf{y})\,\mu(d\mathbf{y})$. The probability of accepting the proposed value $\mathbf{Y}_{t+1}$ as the new value for the chain is $\alpha(\mathbf{X}_t, \mathbf{Y}_{t+1})$, where

$$\alpha(\mathbf{x}, \mathbf{y}) = \begin{cases} \min\left(1, \frac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}\right), & \pi(\mathbf{x})\,q(\mathbf{x},\mathbf{y}) > 0 \\ 1, & \pi(\mathbf{x})\,q(\mathbf{x},\mathbf{y}) = 0 \end{cases}.$$

If the proposed move is accepted, the chain jumps to $\mathbf{X}_{t+1} = \mathbf{Y}_{t+1}$; otherwise, it stays where it is and $\mathbf{X}_{t+1} = \mathbf{X}_t$.

The density q is arbitrary, subject to the condition that the transition kernel be irreducible and aperiodic. The acceptance probability $\alpha(\mathbf{x}, \mathbf{y})$ being chosen to ensure that the chain is reversible with respect to $\pi(\mathbf{y})\,\mu(d\mathbf{y})$, it then follows that the target distribution is stationary for the chain and that the generated Markov chain converges to its stationary distribution.

In this work, the proposed moves are taken to be normally distributed around $\mathbf{x}$, that is $\mathbf{Y}_{t+1} \sim N(\mathbf{X}_t, \sigma^2 I_{d\times d})$ for some σ^2 and with $I_{d\times d}$ the d-dimensional identity matrix. An advantage of this proposal is that it uses the current state of the chain to propose a new value, but yet the proposed value is easily generated. Furthermore, the acceptance probability reduces to

$$\alpha(\mathbf{x}, \mathbf{y}) = \begin{cases} \min\left(1, \frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right), & \pi(\mathbf{x})\,q(\mathbf{x},\mathbf{y}) > 0 \\ 1, & \pi(\mathbf{x})\,q(\mathbf{x},\mathbf{y}) = 0 \end{cases}.$$

In order to have some level of optimality in the performance of the algorithm, care must be exercised when choosing σ^2. If it is too small, the proposed jumps will be too short and therefore simulation will move very slowly to the target distribution in spite of the fact that the proposed moves will be almost all accepted. At the opposite, a large scaling value will generate jumps in low target density regions, resulting in the rejection of the proposed moves and in a chain that stands still most of the time.

To find an appropriate value for σ^2 we need to define a criterion by which measuring efficiency. Roberts et al. in [14] introduce the notion of π-average acceptance rate, which is defined by

$$\int\int \pi(\mathbf{x})\,\alpha(\mathbf{x},\mathbf{y})\,q(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y} \quad = \quad \mathrm{E}\left[1 \wedge \frac{\pi(\mathbf{Y})}{\pi(\mathbf{X})}\right] \tag{1}$$

for the d-dimensional symmetric RWM algorithm. We shall see that this is closely connected to the asymptotic efficiency of the algorithm.

3 The Target Distribution

Consider the following d-dimensional target density

$$\pi\left(d, \mathbf{x}^{(d)}\right) = \prod_{j=1}^{d} \theta_j\left(d\right) f\left(\theta_j\left(d\right) x_j\right). \tag{2}$$

In what follow, we shall refer to $\theta_j^{-2}(d)$, $j = 1, \ldots, d$ as the scaling terms of the target distribution.

We impose the following regularity conditions on the density f: f is a positive C^2 function, $(\log f(X))'$ is Lipschitz continuous,

$$\mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^4\right] = \int_{\mathbf{R}} \left(\frac{f'(x)}{f(x)}\right)^4 f(x)\, dx < \infty,$$

and

$$\mathrm{E}\left[\left(\frac{f''(X)}{f(X)}\right)^2\right] = \int_{\mathbf{R}} \left(\frac{f''(x)}{f(x)}\right)^2 f(x)\, dx < \infty.$$

The product form of the density implies that the d components are independent. They are however not identically distributed, as ascertained by the $\theta_j(d)$'s. In particular, we consider the case where the scaling terms $\theta_j^{-2}(d)$ take the following form

$$\mathbf{\Theta}^{-2}(d) = \left(\frac{K_1}{d^{\lambda_1}}, \ldots, \frac{K_n}{d^{\lambda_n}}, \underbrace{\frac{K_{n+1}}{d^{\gamma_1}}, \ldots, \frac{K_{n+1}}{d^{\gamma_1}}}_{c(\mathcal{J}(1,d))}, \ldots, \underbrace{\frac{K_{n+m}}{d^{\gamma_m}}, \ldots, \frac{K_{n+m}}{d^{\gamma_m}}}_{c(\mathcal{J}(m,d))}\right). \tag{3}$$

That is, some of the terms appear only a fixed number of times while the repetition of others grows with the dimension. Ultimately, we shall be interested in the limit of the target distribution as $d \to \infty$. There is then a need to separate the scaling terms in (3) that will appear only a finite number of times from those that will appear infinitely often as the dimension increases.

More specifically, let $n < \infty$ denote the number of components whose scaling term appears a fixed number of times in $\mathbf{\Theta}^{-2}(d)$. Also, let the j-th of these n scaling terms be K_j/d^{λ_j}, $j = 1, \ldots, n$, where K_j is some positive and finite constant.

Similarly, let $0 < m < \infty$ denote the number of different scaling terms appearing infinitely often in the limit. These m scaling terms are taken to be K_{n+i}/d^{γ_i}, $i = 1, \ldots, m$. For now, we assume that the constants $0 < K_{i+n} < \infty$ are the same for all scaling terms within each of the m groups. We shall relax this assumption in Section 6.

Without loss of generality, we assume the first n and the last $d-n$ scaling terms to be respectively arranged according to an asymptotic increasing order. If $\preceq$ means "is asymptotically smaller than", then we have $\theta_1^{-2}(d) \preceq \ldots \preceq \theta_n^{-2}(d)$ and similarly $\theta_{n+1}^{-2}(d) \preceq \ldots \preceq \theta_d^{-2}(d)$, which respectively implies $-\infty < \lambda_n \leq \lambda_{n-1} \leq \ldots \leq \lambda_1 < \infty$ and $-\infty < \gamma_m \leq \gamma_{m-1} \leq \ldots \leq \gamma_1 < \infty$. In particular, this means that we usually do not consider the constant terms in the determination of this increasing order unless two components have the same power of d, in which case we refer to their constant to determine which term is smaller. Note that with the present notation, two of the first n scaling terms might be identical, in which case we still refer to them as K_j/d^{λ_j} and $K_{j+1}/d^{\lambda_{j+1}}$ with $K_j = K_{j+1}$ and $\lambda_j = \lambda_{j+1}$.

According to this ordering, we can easily determine the asymptotically smallest scaling term $\hat{\theta}^{-2}(d)$, which obviously has to be either the first or the $(n+1)$-st one

$$\hat{\theta}^{-2}(d) = \begin{cases} K_1/d^{\lambda_1}, & \text{if } \lim_{d\to\infty} \frac{K_1/d^{\lambda_1}}{K_{n+1}/d^{\gamma_1}} = 0 \\ K_{n+1}/d^{\gamma_1}, & \text{if } \lim_{d\to\infty} \frac{K_1/d^{\lambda_1}}{K_{n+1}/d^{\gamma_1}} \text{ diverges} \\ \min\left(K_1/d^{\lambda_1}, K_{n+1}/d^{\gamma_1}\right), & \text{if } \lim_{d\to\infty} \frac{K_1/d^{\lambda_1}}{K_{n+1}/d^{\gamma_1}} = K_1/K_{n+1} \end{cases}. \quad (4)$$

To easily refer to the different groups of components whose scaling term appears infinitely often, we define the sets

$$\mathcal{J}(i,d) = \left\{ j \in \{1,\ldots,d\}; \theta_j^{-2}(d) = \frac{K_{i+n}}{d^{\gamma_i}} \right\}$$

for $i = 1,\ldots,m$. The i-th set thus contains positions of components with a scaling term equal to K_{i+n}/d^{γ_i}. These sets are mutually exclusive and their union satisfies $\dot{\bigcup}_{i=1}^{m} \mathcal{J}(i,d) = \{n+1,\ldots,d\}$. We can then write the d-dimensional product density in (2) as

$$\pi\left(d,\mathbf{x}^{(d)}\right) = \prod_{j=1}^{n} \left(\frac{d^{\lambda_j}}{K_j}\right)^{1/2} f\left(\left(\frac{d^{\lambda_j}}{K_j}\right)^{1/2} x_j\right) \prod_{i=1}^{m} \prod_{j\in\mathcal{J}(i,d)} \left(\frac{d^{\gamma_i}}{K_{n+i}}\right)^{1/2} f\left(\left(\frac{d^{\gamma_i}}{K_{n+i}}\right)^{1/2} x_j\right).$$

It is also important to define the cardinality of the sets $\mathcal{J}(i,d)$ since each of the m groups of scaling terms might occupy different proportions of the vector in (3). For $i = 1,\cdots,m$,

$$c(\mathcal{J}(i,d)) = d - n - \sum_{j=1,j\neq i}^{m} c(\mathcal{J}(j,d)) = \#\left\{ j \in \{1,\ldots,d\}; \theta_j^{-2}(d) = \frac{K_{i+n}}{d^{\gamma_i}} \right\}, \quad (5)$$

where $c(\mathcal{J}(i,d))$ is some polynomial function of the dimension satisfying $\lim_{d\to\infty} c(\mathcal{J}(i,d)) = \infty$ and subject to the constraint that the total number of components in the target is d.

To be able to study every component and avoid that some groups of components be undefined as $d \to \infty$, we rearrange the scaling terms in (3):

$$\mathbf{\Theta}^{-2}(d) = \left(\frac{K_1}{d^{\lambda_1}},\ldots,\frac{K_n}{d^{\lambda_n}},\frac{K_{n+1}}{d^{\gamma_1}},\ldots,\frac{K_{n+m}}{d^{\gamma_m}},\frac{K_{n+1}}{d^{\gamma_1}},\ldots,\frac{K_{n+m}}{d^{\gamma_m}},\ldots,\frac{K_{n+1}}{d^{\gamma_1}},\ldots,\frac{K_{n+m}}{d^{\gamma_m}}\right). \quad (6)$$

That is, we take one scaling term from each of the m groups and we put them behind the n scaling terms appearing a fixed number of times. This helps to clearly identify each component being studied as $d \to \infty$ without referring to a component that would otherwise be at an infinite position. Afterwards, we cycle through the components belonging to the different groups of scaling terms, in exercising some caution to preserve the proportion occupied by each of the m groups. This avoids, when $d \to \infty$, assigning infinite positions to the components belonging to the last $m-1$ groups whose scaling term appears infinitely often.

Our goal is to study the limiting distribution of the process for each individual component of the target distribution. To this end, we have to set the scaling term of the component of interest equal to 1. This can be done without loss of generality by applying a linear transformation to the target distribution. This operation is necessary to the obtention of a nontrivial limit for the process involving the component of interest.

4 The Proposal Distribution and its Scaling

A crucial step in the implementation of RWM algorithms is the determination of the optimal form for the proposal scaling as a function of d. Intuitively it would make sense that $\sigma^2(d)$ somehow depends on $\hat{\theta}^{-2}(d)$, the asymptotically smallest scaling term in $\mathbf{\Theta}^{-2}(d)$. Otherwise, the proposed moves might be too large for the components with smaller scaling terms, resulting in a high rejection rate and compromising the convergence of the algorithm. To get a clearer picture of this situation, we can imagine the case where f is the density of the standard normal distribution, in which case $\theta_j^{-2}(d)$ becomes the variance of the j-th component.

Moreover, as the dimension increases the target has more and more components having the same scaling term. Since there is a larger number of individual moves proposed in a single step, it is thus more likely to generate an improbable move for one of the components. To rectify the situation, it is recommended to decrease the proposal scaling as a function of the dimension.

The optimal form for the scaling of the proposal distribution turns out to be $\sigma^2(d) = \ell^2/d^\alpha$, where ℓ^2 is some constant and α is the smallest number satisfying

$$\lim_{d\to\infty} \frac{d^{\lambda_1}}{d^\alpha} < \infty \qquad \text{and} \qquad \lim_{d\to\infty} \frac{d^{\gamma_i} c\left(\mathcal{J}(i,d)\right)}{d^\alpha} < \infty, \quad \text{for } i = 1, \ldots, m. \tag{7}$$

Therefore, at least one of these $m+1$ limits will converge to some positive constant, while the other ones will converge to 0. Since the scaling term of the component of interest is taken equal to one, this implies that the largest possible asymptotical form for the proposal scaling is $\sigma^2 = \sigma^2(d) = \ell^2$, and hence it will never diverge as the dimension grows.

The rationale behind the consideration of all $d-n$ last scaling terms in the determination of α is related to the proportion they occupy in $\mathbf{\Theta}^{-2}(d)$, that is to their corresponding cardinality function $c(\mathcal{J}(i,d))$, $i = 1, \ldots, m$. Indeed, it might turn out that one of these m groups appears in a big enough proportion so as to have more impact than the asymptotically smallest scaling term. In other words, it is possible that the difference in the proportions of two of these groups asymptotically exceeds the difference in the size of the scaling terms.

Having found the optimal form for the scaling of the proposal distribution, we can thus write

$$\mathbf{Y}^{(d)} - \mathbf{x}^{(d)} \sim N\left(\mathbf{0}, \frac{\ell^2}{d^\alpha}\ I_{d\times d}\right).$$

By its nature, the RWM algorithm is a discrete-time process. Since space (the proposal scaling) is function of the dimension of the target distribution, we also have to rescale the time between each step in order to get a nontrivial limiting process as $d \to \infty$. We can make a parallel between our case and Brownian motion expressed as the limit of a simple symmetric random walk. Since we rescaled space through the factor $d^{-\alpha/2}$ (the proposal standard deviation), we have to compensate by speeding up time by a factor of d^α.

Let $\mathbf{Z}^{(d)}(t)$ be the time-t value of the RWM process sped up by a factor of d^α. In particular,

$$\mathbf{Z}^{(d)}(t) = \mathbf{X}^{(d)}([d^\alpha t]) = \left(X_1^{(d)}([d^\alpha t]), \ldots, X_d^{(d)}([d^\alpha t])\right),$$

where $[\cdot]$ is the integer part function. Instead of proposing only one move, the sped up process has the possibility to move on average d^α times during each time interval.

We are now ready to study the limiting comportment of every component in the sequence of processes $\left\{\mathbf{Z}^{(d)}(t), t \geq 0\right\}$ as $d \to \infty$. That is, for each d-dimensional process $\left\{\mathbf{Z}^{(d)}(t), t \geq 0\right\}$ we choose a particular component and study the limiting comportment of this sequence of processes as the dimension increases.

5 Optimal Value for ℓ

We shall now present explicit asymptotic results allowing to determine sensible values for ℓ^2, the constant term of the proposal scaling. We first introduce a weak convergence result for the process $\left\{\mathbf{Z}^{(d)}(t), t \geq 0\right\}$ and most importantly in practice, we transform the conclusion achieved in a statement about efficiency as a function of acceptance rate, as was done in [14].

We denote weak convergence in the Skorokhod topology by $\Rightarrow$, standard Brownian motion at time t by $B(t)$, and the standard normal cumulative distribution function (*cdf*) by $\Phi(\cdot)$. Moreover, recall that the scaling term of the component of interest X_{i^*} is taken to be one, which might require a linear transformation on the target distribution.

Theorem 1. *Consider a RWM algorithm with proposal distribution*

$$\mathbf{Y}^{(d)} \sim N\left(\mathbf{x}^{(d)}, \frac{\ell^2}{d^\alpha} I_{d\times d}\right),$$

where α satisfies (7), and applied to a target density as in (2) satisfying the specified conditions on f, with $\theta_j^{-2}(d)$, $j = 1, \ldots, d$ as in (6). Consider the i^*-th component of the process $\left\{\mathbf{Z}^{(d)}(t), t \geq 0\right\}$, that is $\left\{Z_{i^*}^{(d)}(t), t \geq 0\right\} = \left\{X_{i^*}^{(d)}([d^\alpha t]), t \geq 0\right\}$, and let $\mathbf{X}^{(d)}(0)$ be distributed according to the target density π in (2).

We have

$$\left\{Z_{i^*}^{(d)}(t), t \geq 0\right\} \Rightarrow \{Z(t), t \geq 0\},$$

where $Z(0)$ is distributed according to the density f and $\{Z(t), t \geq 0\}$ satisfies the Langevin stochastic differential equation (SDE)

$$dZ(t) = \upsilon(\ell)^{1/2}\, dB(t) + \frac{1}{2}\upsilon(\ell)\left(\log f(Z(t))\right)' dt,$$

if and only if

$$\lim_{d\to\infty} \frac{\theta_1^2(d)}{\sum_{j=1}^d \theta_j^2(d)} = 0. \tag{8}$$

Here,

$$\upsilon(\ell) = 2\ell^2 \Phi\left(-\frac{\ell\sqrt{E_R}}{2}\right),$$

and

$$E_R = \lim_{d\to\infty} \sum_{i=1}^{m} \frac{c(\mathcal{J}(i,d))}{d^\alpha} \frac{d^{\gamma_i}}{K_{n+i}} \mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^2\right], \tag{9}$$

with $c(\mathcal{J}(i,d))$ as in (5).

Intuitively, we might say that when there is no component converging significantly faster than the others, the limiting process is the same as that found in [14]. In other words, this happens when none of the components possesses a scaling term significantly smaller than the scaling terms of the other components. What we really want is in fact

$$\lim_{d\to\infty} \frac{\hat{\theta}^2(d)}{\sum_{j=1}^d \theta_j^2(d)} = 0,$$

with the reciprocal of $\hat{\theta}^2(d)$ as in (4). However, in the case where $\theta_{n+1}^{-2}(d) \preceq \theta_1^{-2}(d)$, the previous condition is automatically satisfied since in the limit $\theta_{n+1}(d)$ is added an infinite number of times at the denominator, yielding

$$\lim_{d\to\infty} \frac{\theta_1^2(d)}{\sum_{j=1}^d \theta_j^2(d)} \leq \lim_{d\to\infty} \frac{\theta_{n+1}^2(d)}{\sum_{j=1}^d \theta_j^2(d)} = 0.$$

The only case where this condition might be violated is thus when $\theta_1(d) \prec \theta_{n+1}(d)$, from where the importance of Condition (8).

Here $\upsilon(\ell)$ is sometimes interpreted as the speed measure of the diffusion process. This means the limiting process can be expressed as a sped up version of $\{U(t), t \geq 0\}$, a Langevin diffusion process with unity speed measure:

$$\{Z(t), t \geq 0\} = \{U(\upsilon(\ell)t), t \geq 0\},$$

where

$$dU(t) = dB(t) + \frac{1}{2}(\log f(U(t)))' dt.$$

In fact, letting $s = \upsilon(\ell)t$ gives $ds = \upsilon(\ell)dt$ and thus

$$\begin{aligned} dU(s) &= (ds)^{1/2} + \frac{1}{2}\frac{d}{dU(s)}\log f(U(s))\,ds \\ &= (\upsilon(\ell)\,dt)^{1/2} + \frac{1}{2}\frac{d}{dU(\upsilon(\ell)t)}\log f(U(\upsilon(\ell)t))\,\upsilon(\ell)\,dt \\ &= (\upsilon(\ell))^{1/2}\,dB(t) + \frac{1}{2}\upsilon(\ell)\frac{d}{dZ(t)}\log f(Z(t))\,dt \\ &= dZ(t). \end{aligned}$$

The speed measure of the diffusion being proportional to the mixing rate of the algorithm, it suffices to maximize the function $\upsilon(\ell)$ in order to optimize the efficiency of the algorithm.

Let $a(d,\ell)$ be the π-average acceptance rate defined in (1), but where the dependence on the dimension and the proposal scaling are now made explicit. The following corollary introduces the value of ℓ maximizing the speed measure, and thus the efficiency of the RWM algorithm. It also presents the asymptotically optimal acceptance rate, which is of great use for applications.

Corollary 2. *In the settings of Theorem 1 we have* $\lim_{d\to\infty} a(d,\ell) = a(\ell)$, *where*

$$a(\ell) = 2\Phi\left(-\frac{\ell\sqrt{E_R}}{2}\right).$$

Furthermore, $\upsilon(\ell)$ *is maximized at the unique value* $\hat{\ell} = 2.38/\sqrt{E_R}$ *for which* $a\left(\hat{\ell}\right) = 0.234$ *(to three decimal places).*

It is possible to give a simple interpretation to these results. Consider a high-dimensional target distribution as defined in Section 3 to which is applied the RWM algorithm defined in Sections 2 and 4. If there is no component converging significantly faster than the others, the value ℓ should be chosen such that the acceptance rate is close to 0.234 in order to optimize the efficiency of the algorithm. If it is realized that the acceptance rate is substantially larger or smaller than 0.234, the value of $\sigma^2(d)$ should be modified accordingly.

The results presented in this section provide a necessary and sufficient condition to determine in which cases the well-known acceptance rate 0.234 is asymptotically optimal for the target distribution described in Section 3. In particular, these results can be applied to the case where $f(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, yielding a multivariate normal target distribution with independent components. In that case, note that the scaling terms in (6) represent the variances of the individual components. The drift and volatility terms of the limiting Langevin diffusion thus become $-Z(t)/2$ and 1 respectively, and the expression for E_R in (9) can be simplified since $\mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^2\right] = 1$.

More interestingly however, the conclusions of Theorem 1 are also valid for any multivariate normal distribution with correlated components. In fact, since normal random variables are invariant under orthogonal transformations we can transform their covariance matrix in a diagonal matrix where the eigenvalues of the covariance matrix constitute the diagonal elements. The eigenvalues can then be used to verify if Condition (8) is satisfied, and hence to determine whether or not $2.38/\sqrt{E_R}$ is the optimal scaling for the proposal distribution. For example, consider a nontrivial covariance matrix where the variance of each component is 2 and where each covariance term is equal to 1. The d eigenvalues of this matrix are $(d, 1, \ldots, 1)$ and clearly satisfy Condition (8). For a relatively high-dimensional multivariate normal with such a correlation structure, it is thus optimal to tune the acceptance rate to 0.234.

Theorem 1 can also be used to determine if 0.234 is optimal for any normal hierarchical model, since such models possess a distribution which is jointly normal. Consider for instance the simple model where $X_1 \sim N(0,1)$ and $X_j \sim N(X_1, 1)$ for $j = 2, \ldots, d$. The joint distribution of $\mathbf{X}^{(d)}$ is multivariate normal with mean 0 and $d \times d$ covariance matrix such that $\sigma_1^2 = 1$, $\sigma_2^2 = \ldots = \sigma_d^2 = 2$ and $\sigma_{jk}^2 = 1$, $\forall j \neq k$. Using the d eigenvalues, which are $O(d)$, $O(1/d)$ and 1 with multiplicity $d-2$, we thus conclude that Condition (8) is violated and that 0.234 is not optimal, even though the distribution is normal. It is worth mentioning that when the eigenvalues of the covariance matrix take a form different from that assumed in (6) as is the case now, we can apply the more general Theorem 5 in Section 6 instead of Theorem 1.

The previous example might seem surprising as multivariate normal distributions have long been believed to behave as *iid* target distributions in the limit. A natural question to ask is then, what happens when Condition (8) is not satisfied? In such a case, the algorithm can be shown to admit the same limiting Langevin diffusion process but with a different speed measure. Furthermore, the asymptotically optimal acceptance rate is found to be smaller than the usual 0.234. For more details on this case, see [1]. For a better picture of the applicability of these results, examples and simulation studies for various statistical models are presented in [2].

6 Inhomogeneous Proposal Scaling and Extensions

So far, we have assumed the proposal scaling $\sigma^2(d) = \ell^2/d^\alpha$ to be the same for all d components. It is natural to wonder if adjusting the proposal scaling as a function of d for each component would yield a better performance of the algorithm. An important point to keep in mind is that for $\left\{\mathbf{Z}^{(d)}(t), t \geq 0\right\}$ to be a stochastic process, we must speed up time by the same factor for every component. Otherwise, we would face a situation where some components move more frequently than others in the same time interval, and since the acceptance probability of the proposed moves depends on all d components this would violate the definition of a stochastic process. Since d^α is the only time factor yielding a nontrivial limit as $d \to \infty$, we must then keep this parameter fixed. Consequently, the proposal scaling of the component of interest must be ℓ^2/d^α in order to get a nontrivial limiting distribution. Given that we want to study each of the first $n+m$ components, we can thus personalize the proposal scaling of the last $d-n-m$ terms only.

In particular, consider the $\theta_j(d)$'s appearing in (6) and let $\mathbf{Z}^{(d)}(t) = \mathbf{X}^{(d)}([d^\alpha t])$ as before. We set the proposal scaling as follows: for $j = 1, \ldots, n+m$ let $\sigma^2(d) = \ell^2/d^\alpha$ and for $j = n+m+1, \ldots, d$, $j \in \mathcal{J}(i,d)$, let $\sigma^2(d) = \ell^2/\left(c\left(\mathcal{J}(i,d)\right) d^{\gamma_i}\right)$. We have the following result.

Theorem 3. *In the settings of Theorem 1 but with the proposal scaling as just described, the conclusions of Theorem 1 and Corollary 2 are preserved.*

Since the scaling is now adjusted every constant term $K_{n+1}, \ldots, K_{n+m}$ has an impact on the limiting process, yielding a larger value for E_R. Hence, the optimal value $\hat{\ell} = 2.38/\sqrt{E_R}$ is smaller than with homogeneous proposal scaling. When the proposal scaling of all components was based on α, the algorithm had to compensate for the fact that α is chosen as small as possible, and thus maybe too small for certain groups of components, with a larger value for ℓ^2. Since the scaling is now personalized, a smaller value for $\hat{\ell}$ is more appropriate. We note that of course, if the inhomogeneous proposal scaling is identical to the homogeneous one, then both methods will yield the same results and so it is pointless to take about inhomogeneity.

It is also important to see how the conclusions of Section 5 extend to more general target distribution settings than those considered in Section 3. First, we can relax the assumption of equality among the scaling terms $\theta_j^{-2}(d)$ for $j \in \mathcal{J}(i,d)$. That is, we assume the constant terms within each of the m groups to be random and come from some distribution satisfying certain moment conditions. In particular, let

$$\mathbf{\Theta}^{-2}(d) = \left(\frac{K_1}{d^{\lambda_1}}, \ldots, \frac{K_n}{d^{\lambda_n}}, \frac{K_{n+1}}{d^{\gamma_1}}, \ldots, \frac{K_{n+c(\mathcal{J}(1,d))}}{d^{\gamma_1}}, \ldots, \frac{K_{n+\sum_{i=1}^{m-1} c(\mathcal{J}(i,d))+1}}{d^{\gamma_m}}, \ldots, \frac{K_d}{d^{\gamma_m}}\right). \quad (10)$$

We assume that $\{K_j, j \in \mathcal{J}(i,d)\}$ are *iid* and chosen randomly from some distribution with $\mathrm{E}\left[K_j^{-2}\right] < \infty$. Without loss of generality, we also take $\mathrm{E}\left[K_j^{-1/2}\right] = 1$ and denote

$\mathrm{E}\left[K_j^{-1}\right] = b_i$ for $j \in \mathcal{J}(i,d)$. Recall that the scaling term of the component of interest does not depend on d, and we therefore have $\theta_{i^*}^{-2}(d) = K_{i^*}$.

To support the previous modifications, we now suppose that $-\infty < \gamma_m < \gamma_{m-1} < \ldots < \gamma_1 < \infty$. In addition, we suppose that there does not exist a λ_j, $j = 1, \ldots, n$ equal to one of the γ_i, $i = 1, \ldots, m$. This means that if there is an infinite number of scaling terms with the same power of d, they must necessarily belong to the same of the m groups. We obtain the following result.

Theorem 4. *Consider the settings of Theorem 1 with $\mathbf{\Theta}^{-2}(d)$ as in (10) and $\theta_{i^*} = K_{i^*}^{-1/2}$. We have*

$$\left\{Z_{i^*}^{(d)}(t), t \geq 0\right\} \Rightarrow \{Z(t), t \geq 0\},$$

where $Z(0)$ is distributed according to the density $\theta_{i^*} f(\theta_{i^*} x)$ and $\{Z(t), t \geq 0\}$ satisfies the Langevin SDE

$$dZ(t) = (\upsilon(\ell))^{1/2}\, dB(t) + \frac{1}{2}\upsilon(\ell)\left(\log f(\theta_{i^*} Z(t))\right)' dt,$$

if and only if

$$\lim_{d\to\infty} \frac{d^{\lambda_1}}{\sum_{j=1}^{n} d^{\lambda_j} + \sum_{i=1}^{m} c(\mathcal{J}(i,d))\, d^{\gamma_i}} = 0. \tag{11}$$

Here, $\upsilon(\ell)$ is as in Theorem 1 and

$$E_R = \lim_{d\to\infty} \sum_{i=1}^{m} \frac{c(\mathcal{J}(i,d))\, d^{\gamma_i}}{d^{\alpha}}\; b_i \mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^2\right],$$

with

$$c(\mathcal{J}(i,d)) = \#\left\{j \in \{n+1, \ldots, d\}; \theta_j(d) \text{ is } O\left(d^{\gamma_i/2}\right)\right\}.$$

Furthermore, the conclusions of Corollary 2 are preserved.

It is important to notice that Conditions (8) and (11) are equivalent since the constant terms are assumed to be finite. Condition (11) is however easier to verify in the present case due to the randomness of the constant terms.

The previous results can also be extended to more general functions $c(\mathcal{J}(i,d))$, $i = 1, \ldots, m$ and $\theta_j(d)$, $j = 1, \ldots, d$. In order to have sensible limiting theory, we however restrict our attention to functions for which the limit exists as $d \to \infty$. As before, we must also have $c(\mathcal{J}(i,d)) \to \infty$ as $d \to \infty$. We can even allow the scaling terms $\left\{\theta_j^{-2}(d), j \in \mathcal{J}(i,d)\right\}$ to vary within each of the m groups, as long as they are of the same order. That is, for $j \in \mathcal{J}(i,d)$ we suppose

$$\lim_{d\to\infty} \frac{\theta_j(d)}{\theta_i'(d)} = K_j^{-1/2},$$

for some reference function $\theta_i'(d)$ having no constant term (i.e. a constant term equal to 1) and some constant K_j coming from the distribution described for Theorem 4. Note that the reference functions are taken to be as simple as possible. For instance, if all components in a given group $i \in \{1, \ldots, m\}$ are $O(d)$, then $\theta_i'(d) = d$ and not $d+1$.

As for Theorem 4, we assume that if there is infinitely many scaling terms of a certain order they must all belong to one of the m groups. Hence, $\boldsymbol{\Theta}^{-2}(d)$ contains at least m and at most $n+m$ functions of different order. The positions of the elements belonging to the i-th group for $i \in \{1, \ldots, m\}$ are thus given by

$$\mathcal{J}(i,d) = \left\{ j \in \{1, \ldots, d\} ; 0 < \lim_{d \to \infty} \theta_j^{-2}(d)\, \theta_i'^2(d) < \infty \right\}. \tag{12}$$

We again suppose that the scaling terms are classified according to an asymptotic increasing order. In particular, the first n terms of $\boldsymbol{\Theta}^{-2}(d)$ satisfy $\theta_1^{-2}(d) \prec \ldots \prec \theta_n^{-2}(d)$ and the order of the following m terms is chosen to satisfy $\theta_1'^{-2}(d) \prec \ldots \prec \theta_m'^{-2}(d)$.

For such target distributions we define the proposal scaling to be $\sigma^2(d) = \ell^2 \sigma_\alpha^2(d)$, with $\sigma_\alpha^2(d)$ the function of largest possible order such that

$$\lim_{d \to \infty} \theta_1^2(d)\, \sigma_\alpha^2(d) < \infty \quad \text{and} \quad \lim_{d \to \infty} c(\mathcal{J}(i,d))\, \theta_i'^2(d)\, \sigma_\alpha^2(d) < \infty \ \text{ for } i = 1, \ldots, m. \tag{13}$$

We then have the following result.

Theorem 5. *Under the settings of Theorem 4, but with proposal scaling $\sigma^2(d) = \ell^2 \sigma_\alpha^2(d)$ where $\sigma_\alpha^2(d)$ satisfies (13) and with general functions for $c(\mathcal{J}(i,d))$ and $\theta_j(d)$ as defined previously, the conclusions of Theorem 4 are preserved, provided that*

$$\lim_{d \to \infty} \frac{\theta_1^2(d)}{\sum_{j=1}^n \theta_j^2(d) + \sum_{i=1}^m c(\mathcal{J}(i,d))\, \theta_i'^2(d)} = 0$$

holds instead of Condition (8) and with

$$E_R = \lim_{d \to \infty} \sum_{i=1}^m c(\mathcal{J}(i,d))\, \theta_i'^2(d)\, \sigma_\alpha^2(d)\, b_i \mathrm{E}\left[\left(\frac{f'(X)}{f(X)} \right)^2 \right],$$

where $c(\mathcal{J}(i,d))$ is the cardinality function of (12).

This theorem assumes quite a general form for the target distribution and allows for a lot of flexibility. Interestingly, the asymptotically optimal acceptance rate can be shown to be 0.234 as before. For examples such as those discussed at the end of Section 5, Theorem 1 cannot always be applied due to the form of the eigenvalues; this is thus where the general form of Theorem 5 becomes important.

7 Theorems Proofs

We now present the proof of Theorem 1 in Section 5. The proofs of the theorems in Section 6 are similar, so we shall just outline the main differences at the end of this section.

The proof of Theorem 1 is based on Corollary 8.1 of Chapter 4 in [9]. This corollary roughly says that for the finite-dimensional distributions of a sequence of processes to converge weakly to those of some Markov process, it is enough to verify $\mathcal{L}^1$ convergence of their generators. To reach weak convergence of the stochastic processes themselves, Theorem 7.8 of Chapter 3 in [9] states that it is sufficient to verify relative compactness for each stochastic process in the sequence. This step is easily checked using Corollary 7.4 and Theorem 8.6 of Chapter 3 in [9] along with a continuity of probabilities argument and the fact that the algorithm starts in stationarity.

Our task is then to focus on the proof of the convergence in mean of the generators. To this end, we base our approach on the proof for the RWM algorithm case in [13]. Note however that the authors instead prove uniform convergence of generators, which could not be used in the present situation.

The definition of generators is written in term of an arbitrary test function h, which can usually be any smooth function. In the present case, we can however restrict our attention to functions in C_c^∞, the space of infinitely differentiable functions on compact support. Since the limiting process obtained is a diffusion process, then C_c^∞ is a core for the generator by Theorem 2.1 of Chapter 8 in [9]. A core is roughly defined to be representative enough so as to focus on the functions it contains only.

In order to lighten the formulas, we adopt the following convention for defining vectors. The number in parentheses (say a) appearing at the exponent denotes the first a components of the d-dimensional vector. When a substraction of terms appears in the parentheses (say $b-a$), the vector contains the first b components from which we removed the first a, so it is formed of the components $a+1,\ldots,b$. The minus sign appearing outside the brackets informs us that the i^*-th component, i.e. the component of interest, is excluded from the vector. For instance, the vector $\mathbf{X}^{(d-n)-}$ contains the last $d-n$ target random variables, i.e. the components having a scaling term appearing infinitely often and from which we excluded the i^*-th component.

We also adopt the following convention for conditional expectations. The expectation is computed with respect to the variables appearing as a subscript on the right of the operator E. When there is no such subscript, this means that the expectation is taken with respect to all random variables included in the expression. When there is possibility of confusion, we include the subscript even if it is not necessary according to this convention. For instance, we write

$$\mathrm{E}\left[f\left(X,Y\right)\right]=\mathrm{E}\left[\mathrm{E}\left[f\left(X,Y\right)|Y\right]\right]=\mathrm{E}_Y\left[\mathrm{E}_X\left[f\left(X,Y\right)\right]\right].$$

7.1 Restrictions on the Proposal Scaling

The first step of the proof is to transform Condition (8) in a statement about the proposal scaling and its parameter α. Developing the denominator of the condition yields

$$\sum_{j=1}^{d} \theta_j^2(d) = \frac{d^{\lambda_1}}{K_1} + \ldots + \frac{d^{\lambda_n}}{K_n} + c(\mathcal{J}(1,d)) \frac{d^{\gamma_1}}{K_{n+1}} + \ldots + c(\mathcal{J}(m,d)) \frac{d^{\gamma_m}}{K_{n+m}}.$$

In order for the condition to be satisfied, we must equivalently have

$$\begin{aligned} &\lim_{d\to\infty} \sum_{j=1}^{d} \theta_1^{-2}(d)\, \theta_j^2(d) \\ &= \lim_{d\to\infty} \frac{K_1}{d^{\lambda_1}} \left(\frac{d^{\lambda_1}}{K_1} + \ldots + \frac{d^{\lambda_n}}{K_n} + c(\mathcal{J}(1,d)) \frac{d^{\gamma_1}}{K_{n+1}} + \ldots + c(\mathcal{J}(m,d)) \frac{d^{\gamma_m}}{K_{n+m}} \right) = \infty. \end{aligned}$$

Letting $b = \max(j \in \{1,\ldots,n\}; \lambda_j = \lambda_1)$ the number of components with a scaling term of the same order as that of X_1, we obtain

$$\lim_{d\to\infty} \theta_1^{-2}(d) \left(\frac{d^{\lambda_1}}{K_1} + \ldots + \frac{d^{\lambda_n}}{K_n} \right) = 1 + \sum_{j=2}^{b} \frac{K_1}{K_j} < \infty.$$

To have an overall limit that is infinite, it must then be true that

$$\lim_{d\to\infty} \theta_1^{-2}(d) \left(c(\mathcal{J}(1,d)) \frac{d^{\gamma_1}}{K_{n+1}} + \ldots + c(\mathcal{J}(m,d)) \frac{d^{\gamma_m}}{K_{n+m}} \right) = \infty,$$

that is for at least one $i \in \{1,\cdots,m\}$,

$$\lim_{d\to\infty} \theta_1^{-2}(d)\, c(\mathcal{J}(i,d)) \frac{d^{\gamma_i}}{K_{n+i}} = \lim_{d\to\infty} \frac{K_1}{K_{n+i}} \frac{c(\mathcal{J}(i,d))\, d^{\gamma_i}}{d^{\lambda_1}} = \infty. \tag{14}$$

This implies that the form of the scaling of the proposal distribution, i.e. the choice of the parameter α, must be based on one of the groups of scaling terms appearing infinitely often. In other words, it cannot possibly be based on K_1/d^{λ_1}, the smallest scaling term appearing a fixed number of times. If it was, this would mean that

$$\lim_{d\to\infty} \frac{c(\mathcal{J}(i,d))\, d^{\gamma_i}}{d^{\alpha}} = \lim_{d\to\infty} \frac{c(\mathcal{J}(i,d))\, d^{\gamma_i}}{d^{\lambda_1}} = \infty,$$

for all i for which (14) was diverging, which would contradict the definition of α. Therefore when Condition (8) is satisfied, it follows that $\lim_{d\to\infty} d^{\lambda_1}/d^{\alpha} = 0$ and $\theta_1^{-2}(d)$ does not have any impact on the determination of the parameter α. This thus implies that α is always strictly greater than 0, no matter which component is under study.

7.2 Proof of Theorem 1

We now demonstrate that the generator of the RWM algorithm converges in mean to that of the Langevin diffusion. To this end, we shall use the results appearing in Sections 8 and 9.

Proof. We need to show that for an arbitrary test function $h \in C_c^{\infty}$,

$$\lim_{d\to\infty} \mathrm{E}\left[\left|Gh\left(d, X_{i^*}\right) - G_L h\left(X_{i^*}\right)\right|\right] = 0,$$

where

$$Gh\left(d, X_{i^*}\right) = d^{\alpha} \mathrm{E}_{\mathbf{Y}^{(d)}, \mathbf{X}^{(d)-}}\left[\left(h\left(Y_{i^*}\right) - h\left(X_{i^*}\right)\right)\left(1 \wedge \frac{\pi\left(d, \mathbf{Y}^{(d)}\right)}{\pi\left(d, \mathbf{X}^{(d)}\right)}\right)\right]$$

is the discrete-time generator of the sped up Metropolis-Hastings algorithm, and

$$G_L\left(X_{i^*}\right) = \upsilon\left(\ell\right)\left[\frac{1}{2} h''\left(X_{i^*}\right) + \frac{1}{2} h'\left(X_{i^*}\right)\left(\log f\left(X_{i^*}\right)\right)'\right]$$

is the generator of a Langevin diffusion process with speed measure $\upsilon\left(\ell\right)$ as in Theorem 1.

We begin by introducing a third generator $\widetilde{G}h\left(d, X_{i^*}\right)$ (as in (15) of Lemma 7) that is asymptotically equivalent to the original generator $Gh\left(d, X_{i^*}\right)$. By the triangle's inequality,

$$\begin{aligned}
&\mathrm{E}\left[\left|Gh\left(d, X_{i^*}\right) - \widetilde{G}h\left(d, X_{i^*}\right) + \widetilde{G}h\left(d, X_{i^*}\right) - G_L h\left(X_{i^*}\right)\right|\right] \\
&\quad\leq \mathrm{E}\left[\left|Gh\left(d, X_{i^*}\right) - \widetilde{G}h\left(d, X_{i^*}\right)\right|\right] + \mathrm{E}\left[\left|\widetilde{G}h\left(d, X_{i^*}\right) - G_L h\left(X_{i^*}\right)\right|\right].
\end{aligned}$$

From Lemma 7, the first expectation on the RHS converges to 0 as $d \to \infty$. To prove the theorem, we are thus left to show $\mathcal{L}^1$ convergence of the generator $\widetilde{G}h\left(d, X_{i^*}\right)$ to that of the Langevin diffusion.

Substituting explicit expressions for the generators and the speed measure, grouping some terms and using the triangle's inequality yield

$$\begin{aligned}
&\mathrm{E}\left[\left|\widetilde{G}h\left(d, X_{i^*}\right) - G_L h\left(X_{i^*}\right)\right|\right] \\
&\quad\leq \ell^2 \left|\frac{1}{2}\mathrm{E}\left[1 \wedge e^{\sum_{j=1, j\neq i^*}^{d} \varepsilon\left(d, X_j, Y_j\right)}\right] - \Phi\left(-\frac{\ell\sqrt{E_R}}{2}\right)\right| \mathrm{E}\left[\left|h''\left(X_{i^*}\right)\right|\right] \\
&\qquad + \ell^2 \left|\mathrm{E}\left[e^{\sum_{j=1, j\neq i^*}^{d} \varepsilon\left(d, X_j, Y_j\right)}; \sum_{j=1, j\neq i^*}^{d} \varepsilon\left(d, X_j, Y_j\right) < 0\right] - \Phi\left(-\frac{\ell\sqrt{E_R}}{2}\right)\right| \\
&\qquad \times \mathrm{E}\left[\left|h'\left(X_{i^*}\right)\left(\log f\left(X_{i^*}\right)\right)'\right|\right].
\end{aligned}$$

Since the function h has compact support, it implies that h itself and its derivatives are bounded by some constant. Therefore, $\mathrm{E}\left[\left|h''\left(X_{i^*}\right)\right|\right]$ and $\mathrm{E}\left[\left|h'\left(X_{i^*}\right)\left(\log f\left(X_{i^*}\right)\right)'\right|\right]$ are both bounded by K, say. Using Lemmas 8 and 9, we then conclude that the first absolute difference on the RHS goes to 0 as $d \to \infty$, and we reach the same conclusion for the second absolute difference by applying Lemmas 10 and 11. □

7.3 Proof of Theorem 4

Most of the proof is very similar to that of Theorem 1. The main difference happens when working with any one of the m different groups formed of infinitely many components. Since the constant terms are now random, we cannot factorize the scaling terms of components belonging to a same group. This difficulty is however easily overcome by changes of variables and the use of conditional expectations; for instance, a typical situation we face is

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{\Theta}^{(d)}_{\mathcal{J}(i,d)},\mathbf{X}^{(d)}_{\mathcal{J}(i,d)}}\left[\sum_{j\in\mathcal{J}(i,d)}\left(\frac{d}{dX_j}\log\theta_j\left(d\right)f\left(\theta_j\left(d\right)X_j\right)\right)^2\right] \\
&= \mathrm{E}_{\boldsymbol{\Theta}^{(d)}_{\mathcal{J}(i,d)}}\left[\sum_{j\in\mathcal{J}(i,d)}\int\left(\frac{f'\left(\theta_j\left(d\right)x_j\right)}{f\left(\theta_j\left(d\right)x_j\right)}\right)^2\theta_j\left(d\right)f\left(\theta_j\left(d\right)x_j\right)dx_j\right] \\
&= \sum_{j\in\mathcal{J}(i,d)}\mathrm{E}\left[\theta_j^2\left(d\right)\right]\int\left(\frac{f'\left(x\right)}{f\left(x\right)}\right)^2 f\left(x\right)dx \\
&= c\left(\mathcal{J}\left(i,d\right)\right)b_i d^{\gamma_i}\mathrm{E}\left[\left(\frac{f'\left(X\right)}{f\left(X\right)}\right)^2\right],
\end{aligned}$$

where $\mathbf{X}^{(d)}_{\mathcal{J}(i,d)}$ is the vector containing the random variables $\{X_j, j\in\mathcal{J}\left(i,d\right)\}$ and similarly for $\boldsymbol{\Theta}^{(d)}_{\mathcal{J}(i,d)}$. Instead of carrying the term $\theta^2_{n+i}\left(d\right)=d^{\gamma_i}/K_{n+i}$, we thus carry $b_i d^{\gamma_i}$.

7.4 Proof of Theorem 5

The general forms of the functions $c\left(\mathcal{J}\left(i,d\right)\right)$, $i=1,\ldots,m$ and $\theta_j\left(d\right)$, $j=1,\ldots,d$ necessitate a fancier notation, but do not affect the body of the proof. What alters the demonstration is rather the fact that the scaling terms $\theta_j\left(d\right)$ for $j\in\mathcal{J}\left(i,d\right)$ are allowed to be different functions of the dimension as long as they are of the same order. Because of this particularity of the model, we have to write

$$\theta_j\left(d\right)=K_j^{-1/2}\theta'_i\left(d\right)\frac{\theta^*_j\left(d\right)}{\theta'_i\left(d\right)},$$

where $\theta^*_j\left(d\right)$ is implicitly defined. We can then carry with the proof as usual, factoring the term $b_i\theta'_i\left(d\right)$ instead of $\theta^2_{n+i}\left(d\right)$ in Theorem 1 (or $b_i d^{\gamma_i}$ in Theorem 4). Since $\lim_{d\to\infty}\theta^*_j\left(d\right)/\theta'_i\left(d\right)=1$, the rest of the proof can be repeated with minor modifications.

8 Approximate Generator and Other Results

8.1 Convergence of an Approximation Term

The following result shall be of great use in the demonstration of many of the subsequent lemmas, which in turn will be used to prove Theorem 1.

Lemma 6. *For $i = 1, \ldots, m$, let*

$$\begin{aligned} W_i\left(d, \mathbf{X}^{(d)}_{\mathcal{J}(i,d)}, \mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}\right) &= \frac{1}{2}\sum_{j \in \mathcal{J}(i,d)} \left(\frac{d^2}{dX_j^2}\log\theta_j(d) f(\theta_j(d) X_j)\right)(Y_j - X_j)^2 \\ &+ \frac{\ell^2}{2d^\alpha}\sum_{j \in \mathcal{J}(i,d)} \left(\frac{d}{dX_j}\log\theta_j(d) f(\theta_j(d) X_j)\right)^2, \end{aligned}$$

where $Y_j \,|X_j \sim N(X_j, \ell^2/d^\alpha)$ and X_j is distributed according to the density $\theta_j(d) f(\theta_j(d) x_j)$, independently for all $j = 1, ..., d$. Then for $i = 1, \ldots, m$

$$\mathrm{E}\left[\left|W_i\left(d, \mathbf{X}^{(d)}_{\mathcal{J}(i,d)}, \mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}\right)\right|\right] \;\to\; 0 \text{ as } d \to \infty.$$

Proof. By Jensen's inequality

$$\mathrm{E}_{\mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}}\left[\left|W_i\left(d, \mathbf{X}^{(d)}_{\mathcal{J}(i,d)}, \mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}\right)\right|\right] \leq \sqrt{\mathrm{E}_{\mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}}\left[W_i^2\left(d, \mathbf{X}^{(d)}_{\mathcal{J}(i,d)}, \mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}\right)\right]}.$$

Developing the square, taking the expectation conditional on $\mathbf{X}^{(d)}_{\mathcal{J}(i,d)}$ and factoring out $\frac{\ell^4}{4d^{2\alpha}}$ yield

$$\begin{aligned} &\mathrm{E}_{\mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}}\left[W_i^2\left(d, \mathbf{X}^{(d)}_{\mathcal{J}(i,d)}, \mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}\right)\right] = \frac{\ell^4}{4d^{2\alpha}} \times \\ &\left\{3\sum_{j \in \mathcal{J}(i,d)}\left(\frac{d^2}{dX_j^2}\log\theta_j(d) f(\theta_j(d) X_j)\right)^2 + \sum_{j \in \mathcal{J}(i,d)}\left(\frac{d}{dX_j}\log\theta_j(d) f(\theta_j(d) X_j)\right)^4\right. \\ &+2\sum_{k \in \mathcal{J}(i,d)}\sum_{j=\mathcal{J}_{k+1}(i,d)}^{\mathcal{J}_{c(\mathcal{J}(i,d))}(i,d)} \frac{d^2}{dX_j^2}\log\theta_j(d) f(\theta_j(d) X_j)\frac{d^2}{dX_k^2}\log\theta_k(d) f(\theta_k(d) X_k) \\ &+2\sum_{k \in \mathcal{J}(i,d)}\sum_{j=\mathcal{J}_{k+1}(i,d)}^{\mathcal{J}_{c(\mathcal{J}(i,d))}(i,d)} \left(\frac{d}{dX_j}\log\theta_j(d) f(\theta_j(d) X_j)\right)^2\left(\frac{d}{dX_k}\log\theta_k(d) f(\theta_k(d) X_k)\right)^2 \\ &\left.+2\sum_{k \in \mathcal{J}(i,d)}\sum_{j \in \mathcal{J}(i,d)} \frac{d^2}{dX_j^2}\log\theta_j(d) f(\theta_j(d) X_j)\left(\frac{d}{dX_k}\log\theta_k(d) f(\theta_k(d) X_k)\right)^2\right\}. \end{aligned}$$

The previous expression can be reexpressed as

$$\mathrm{E}_{\mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}}\left[W_i^2\left(d, \mathbf{X}^{(d)}_{\mathcal{J}(i,d)}, \mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}\right)\right] = \frac{\ell^4}{2d^{2\alpha}}\sum_{j \in \mathcal{J}(i,d)}\left(\frac{d^2}{dX_j^2}\log\theta_j(d) f(\theta_j(d) X_j)\right)^2$$

$$+\frac{\ell^4}{4d^{2\alpha}}\left\{\sum_{j\in\mathcal{J}(i,d)}\left(\frac{d^2}{dX_j^2}\log\theta_j(d)f(\theta_j(d)X_j)+\left(\frac{d}{dX_j}\log\theta_j(d)f(\theta_j(d)X_j)\right)^2\right)\right\}^2,$$

and hence

$$\sqrt{\mathrm{E}_{\mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}}\left[W_i^2\left(d,\mathbf{X}^{(d)}_{\mathcal{J}(i,d)},\mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}\right)\right]}\leq\frac{\ell^2}{\sqrt{2}d^\alpha}\left(\sum_{j\in\mathcal{J}(i,d)}\left(\frac{d^2}{dX_j^2}\log\theta_j(d)f(\theta_j(d)X_j)\right)^2\right)^{1/2}$$
$$+\frac{\ell^2}{2d^\alpha}\left|\sum_{j\in\mathcal{J}(i,d)}\left(\frac{d^2}{dX_j^2}\log\theta_j(d)f(\theta_j(d)X_j)+\left(\frac{d}{dX_j}\log\theta_j(d)f(\theta_j(d)X_j)\right)^2\right)\right|.$$

Using changes of variables, the unconditional expectation then satisfies

$$\mathrm{E}\left[\left|W_i\left(d,\mathbf{X}^{(d)}_{\mathcal{J}(i,d)},\mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}\right)\right|\right]\leq\frac{\ell^2}{\sqrt{2}d^\alpha}\theta^2_{n+i}(d)\sqrt{c(\mathcal{J}(i,d))}\mathrm{E}\left[\left(\frac{d^2}{dX^2}\log f(X)\right)^2\right]^{1/2}$$
$$+\frac{\ell^2}{2d^\alpha}\theta^2_{n+i}(d)\,c(\mathcal{J}(i,d))\,\mathrm{E}\left[\left|\frac{1}{c(\mathcal{J}(i,d))}\sum_{j\in\mathcal{J}(i,d)}\left(\frac{d^2}{dX_j^2}\log f(X_j)+\left(\frac{d}{dX_j}\log f(X_j)\right)^2\right)\right|\right].$$

Considering that $d^\alpha < d^{\gamma_i}\sqrt{c(\mathcal{J}(i,d))}$ along with the fact that the expectation in the first term on the RHS is bounded by some constant, it implies that this term converges to 0 as $d\to\infty$. It thus remains to verify the convergence of the second term. Since

$$\frac{\ell^2}{2d^\alpha}\frac{d^{\gamma_i}}{K_{n+i}}c(\mathcal{J}(i,d))$$

is $O(1)$ for at least one $i\in\{1,\ldots,m\}$, we must then show that the expectation converges to 0 as the dimension increases. We have

$$\mathrm{E}\left[\frac{d^2}{dX_j^2}\log f(X_j)+\left(\frac{d}{dX_j}\log f(X_j)\right)^2\right]=\mathrm{E}\left[\frac{f''(X)}{f(X)}\right]=\int f''(x)\,dx=0$$

and

$$\mathrm{Var}\left(\frac{f''(X)}{f(X)}\right)=\mathrm{E}\left[\left(\frac{f''(X)}{f(X)}\right)^2\right]<\infty$$

by assumption. By the WLLN,

$$|S_i(d)|\equiv\left|\frac{1}{c(\mathcal{J}(i,d))}\sum_{j\in\mathcal{J}(i,d)}\left(\frac{d^2}{dX_j^2}\log f(X_j)+\left(\frac{d}{dX_j}\log f(X_j)\right)^2\right)\right|\to_p 0\quad\text{as } d\to\infty.$$

We now want to verify if we could bring the limit inside the expectation. By independence between the X_j's, we find

$$\mathrm{E}\left[(S_i(d))^2\right]=\mathrm{E}\left[\left(\frac{1}{c(\mathcal{J}(i,d))}\sum_{j\in\mathcal{J}(i,d)}\frac{f''(X_j)}{f(X_j)}\right)^2\right]=\frac{1}{c(\mathcal{J}(i,d))}\mathrm{E}\left[\left(\frac{f''(X)}{f(X)}\right)^2\right]<\infty$$

for all d. Then,

$$\sup_d \mathrm{E}\left[|S_i(d)|\,\mathbf{1}_{\{|S_i(d)|\geq a\}}\right] \leq \sup_d \frac{1}{a}\mathrm{E}\left[(S_i(d))^2\,\mathbf{1}_{\{|S_i(d)|\geq a\}}\right] \leq \frac{K}{a} \to 0 \quad \text{as } a \to \infty.$$

Since the uniform integrability condition is satisfied (see for instance [5], [10] or [18]), we can bring the limit inside the expectation and find

$$\lim_{d\to\infty} \mathrm{E}\left[|S_i(d)|\right] = \mathrm{E}\left[\lim_{d\to\infty} |S_i(d)|\right] = 0,$$

which completes the proof of the lemma. □

8.2 Convergence to the Approximate Generator $\widetilde{G}h(d, X_{i^*})$

We prove a result stating that the discrete-time generator of the sped up RWM algorithm is asymptotically equivalent to the approximate generator $\widetilde{G}h(d, X_{i^*})$.

Lemma 7. *For any function $h \in C_c^\infty$, let*

$$\begin{aligned}\widetilde{G}h(d, X_{i^*}) &= \frac{1}{2}\ell^2 h''(X_{i^*})\,\mathrm{E}\left[1 \wedge e^{\sum_{j=1,j\neq i^*}^{d} \varepsilon(d,X_j,Y_j)}\right] \\ &\quad + \ell^2 h'(X_{i^*})\,(\log f(X_{i^*}))'\,\mathrm{E}\left[e^{\sum_{j=1,j\neq i^*}^{d} \varepsilon(d,X_j,Y_j)};\ \sum_{j=1,j\neq i^*}^{d} \varepsilon(d, X_j, Y_j) < 0\right],\end{aligned} \tag{15}$$

where

$$\varepsilon(d, X_j, Y_j) = \log \frac{f(\theta_j(d)\,Y_j)}{f(\theta_j(d)\,X_j)}. \tag{16}$$

Then if $\alpha > 0$ as defined in (7),

$$\lim_{d\to\infty} \mathrm{E}\left[\left|Gh(d, X_{i^*}) - \widetilde{G}h(d, X_{i^*})\right|\right] = 0.$$

Proof. The generator of the sped up RWM algorithm is

$$\begin{aligned}Gh(d, X_{i^*}) &= d^\alpha \mathrm{E}_{\mathbf{Y}^{(d)},\mathbf{X}^{(d)-}}\left[(h(Y_{i^*}) - h(X_{i^*}))\left(1 \wedge \frac{\pi\left(d, \mathbf{Y}^{(d)}\right)}{\pi\left(d, \mathbf{X}^{(d)}\right)}\right)\right] \\ &= d^\alpha \mathrm{E}_{Y_{i^*}}\left[(h(Y_{i^*}) - h(X_{i^*}))\,\mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[1 \wedge \frac{\pi\left(d, \mathbf{Y}^{(d)}\right)}{\pi\left(d, \mathbf{X}^{(d)}\right)}\right]\right]\end{aligned}$$

We first concentrate on the inner expectation. Using properties of the log function, we get

$$\begin{aligned}
&\mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[1\wedge\frac{\pi\left(d,\mathbf{Y}^{(d)}\right)}{\pi\left(d,\mathbf{X}^{(d)}\right)}\right] \\
&= \mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[1\wedge\exp\left\{\log\frac{f\left(Y_{i^*}\right)}{f\left(X_{i^*}\right)}+\sum_{j=1,j\neq i^*}^{d}\log\frac{f\left(\theta_j\left(d\right)Y_j\right)}{f\left(\theta_j\left(d\right)X_j\right)}\right\}\right] \\
&= \mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[1\wedge\exp\left\{\varepsilon\left(X_{i^*},Y_{i^*}\right)+\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)\right\}\right],
\end{aligned}$$

where

$$\varepsilon\left(X_{i^*},Y_{i^*}\right)=\log\frac{f\left(Y_{i^*}\right)}{f\left(X_{i^*}\right)}\quad\text{and}\quad\varepsilon\left(d,X_j,Y_j\right)=\log\frac{f\left(\theta_j\left(d\right)Y_j\right)}{f\left(\theta_j\left(d\right)X_j\right)}.$$

We can thus express the generator as

$$Gh\left(d,X_{i^*}\right)=d^{\alpha}\mathrm{E}_{Y_{i^*}}\left[\left(h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\right)\mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[1\wedge e^{\sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)}\right]\right]. \tag{17}$$

We shall compute the outside expectation. To this effect, a Taylor expansion of the minimum function with respect to Y_{i^*} and around X_{i^*} is used. Since f is a C^2 density function, the minimum function will be twice differentiable as well except at the points where $\sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)=0$. This will however not affect the expectation, since the set of values at which the derivatives do not exist has Lebesgue probability 0. The first and second derivatives of the minimum function are

$$\frac{\partial}{\partial Y_{i^*}}1\wedge e^{\sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)} = \begin{cases}\frac{\partial}{\partial Y_{i^*}}\varepsilon\left(X_{i^*},Y_{i^*}\right)e^{\sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)} & \text{if } \sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)<0 \\ 0 & \text{if } \sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)>0\end{cases},$$

and

$$\begin{aligned}
&\frac{\partial^2}{\partial Y_{i^*}^2}1\wedge e^{\sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)} \\
&= \begin{cases}\left(\frac{\partial^2}{\partial Y_{i^*}^2}\varepsilon\left(X_{i^*},Y_{i^*}\right)+\left(\frac{\partial}{\partial Y_{i^*}}\varepsilon\left(X_{i^*},Y_{i^*}\right)\right)^2\right)e^{\sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)} & \text{if } \sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)<0 \\ 0 & \text{if } \sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)>0\end{cases}.
\end{aligned}$$

Expressing the inner expectation in (17) as a function of these derivatives, we find

$$\begin{aligned}
&\mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[1\wedge e^{\sum_{j=1}^{d}\varepsilon\left(d,X_j,Y_j\right)}\right] \\
&= \mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[1\wedge e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)}\right] \\
&\quad+\left(Y_{i^*}-X_{i^*}\right)\left(\log f\left(X_{i^*}\right)\right)'\mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)};\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)<0\right] \\
&\quad+\frac{1}{2}\left(Y_{i^*}-X_{i^*}\right)^2\left(\left(\log f\left(U_{i^*}\right)\right)'+\left(\log f\left(U_{i^*}\right)\right)''\right)\mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[e^{g\left(U_{i^*}\right)};g\left(U_{i^*}\right)<0\right],
\end{aligned}$$

where

$$g\left(U_{i^*}\right)=\varepsilon\left(X_{i^*},U_{i^*}\right)+\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right),$$

for some $U_{i^*}\in\left(X_{i^*},Y_{i^*}\right)$ or $\left(Y_{i^*},X_{i^*}\right)$.

Using this expansion, the generator becomes

$$\begin{aligned}
&Gh\left(d,X_{i^*}\right)\\
&= d^\alpha \mathrm{E}_{Y_{i^*}}\left[\left(h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\right)\right]\mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[1\wedge e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)}\right]\\
&\quad+d^\alpha\left(\log f\left(X_{i^*}\right)\right)'\mathrm{E}_{Y_{i^*}}\left[\left(h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\right)\left(Y_{i^*}-X_{i^*}\right)\right]\\
&\quad\times\mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)};\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)<0\right]\\
&\quad+\frac{d^\alpha}{2}\mathrm{E}_{Y_{i^*}}\left[\left(h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\right)\left(Y_{i^*}-X_{i^*}\right)^2\left(\log f\left(U_{i^*}\right)\right)'\mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[e^{g(U_{i^*})};g\left(U_{i^*}\right)<0\right]\right]\\
&\quad+\frac{d^\alpha}{2}\mathrm{E}_{Y_{i^*}}\left[\left(h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\right)\left(Y_{i^*}-X_{i^*}\right)^2\left(\log f\left(U_{i^*}\right)\right)''\mathrm{E}_{\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}}\left[e^{g(U_{i^*})};g\left(U_{i^*}\right)<0\right]\right].
\end{aligned}$$

Again, a Taylor's expansion yields

$$\begin{aligned}
&h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\\
&= h'\left(X_{i^*}\right)\left(Y_{i^*}-X_{i^*}\right)+\frac{1}{2}h''\left(X_{i^*}\right)\left(Y_{i^*}-X_{i^*}\right)^2+\frac{1}{6}h'''\left(V_{i^*}\right)\left(Y_{i^*}-X_{i^*}\right)^3,
\end{aligned}\tag{18}$$

for some V_{i^*} lying between X_{i^*} and Y_{i^*}. Since the function h has compact support, then h itself and its derivatives are bounded by some positive constant (say K), which gives

$$d^\alpha\mathrm{E}_{Y_{i^*}}\left[h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\right]\le\frac{\ell^2}{2}h''\left(X_{i^*}\right)+\frac{\ell^3}{6}\sqrt{\frac{8}{\pi}}\frac{K}{d^{\alpha/2}},$$

along with

$$d^\alpha\mathrm{E}_{Y_{i^*}}\left[h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\left(Y_{i^*}-X_{i^*}\right)\right]\le\ell^2h'\left(X_{i^*}\right)+\frac{\ell^4}{2d^\alpha}K.$$

Substituting these expressions in the equation for $Gh\left(d,X_{i^*}\right)$, noticing that all expectations computed with respect to $\mathbf{Y}^{(d)-}$ and $\mathbf{X}^{(d)-}$ are bounded by one and that $\left|\left(\log f\left(U_{i^*}\right)\right)''\right|$ is bounded by some positive constant K, we obtain

$$\begin{aligned}
Gh\left(d,X_{i^*}\right)&\le\widetilde{G}h\left(d,X_{i^*}\right)+\frac{\ell^3}{6}\sqrt{\frac{8}{\pi}}\frac{K}{d^{\alpha/2}}+\frac{\ell^4}{2d^\alpha}K\left|\left(\log f\left(X_{i^*}\right)\right)'\right|\\
&\quad+\frac{d^\alpha}{2}\mathrm{E}_{Y_{i^*}}\left[\left|h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\right|\left(Y_{i^*}-X_{i^*}\right)^2\left|\left(\log f\left(U_{i^*}\right)\right)'\right|\right]\\
&\quad+\frac{d^\alpha}{2}K\mathrm{E}_{Y_{i^*}}\left[\left|h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\right|\left(Y_{i^*}-X_{i^*}\right)^2\right].
\end{aligned}\tag{19}$$

Using a two-term Taylor expansion around X_{i^*}, we find

$$\begin{aligned}\left|\left(\log f\left(U_{i^*}\right)\right)'\right| &= \left|\left(\log f\left(X_{i^*}\right)\right)' + \left(\log f\left(V_{i^*}\right)\right)''\left(U_{i^*} - X_{i^*}\right)\right| \\ &\leq \left|\left(\log f\left(X_{i^*}\right)\right)'\right| + K\left|Y_{i^*} - X_{i^*}\right|,\end{aligned}$$

where $V_{i^*} \in (X_{i^*}, U_{i^*})$ or (U_{i^*}, X_{i^*}). In addition, using (18) yields

$$d^{\alpha}\mathrm{E}_{Y_{i^*}}\left[\left|h\left(Y_{i^*}\right) - h\left(X_{i^*}\right)\right|\left(Y_{i^*} - X_{i^*}\right)^2\right] \leq \frac{\ell^3}{d^{\alpha/2}}\sqrt{\frac{8}{\pi}}K + \frac{3\ell^4}{2d^{\alpha}}K + \frac{\ell^5}{d^{3\alpha/2}}\sqrt{\frac{32}{\pi}}\frac{K}{3},$$

and

$$d^{\alpha}\mathrm{E}_{Y_{i^*}}\left[\left|h\left(Y_{i^*}\right) - h\left(X_{i^*}\right)\right|\left|Y_{i^*} - X_{i^*}\right|^3\right] \leq 3\frac{\ell^4}{d^{\alpha}}K + \sqrt{\frac{32}{\pi}}\frac{\ell^5}{d^{3\alpha/2}}K + \frac{5}{2}\frac{\ell^6}{d^{2\alpha}}K.$$

We can then simplify (19) further and write

$$\begin{aligned}&Gh\left(d, X_{i^*}\right) - \widetilde{G}h\left(d, X_{i^*}\right) \\ &\leq \frac{\ell^3}{6}\sqrt{\frac{8}{\pi}}\frac{K}{d^{\alpha/2}} + \frac{\ell^4}{2d^{\alpha}}K\left|\left(\log f\left(X_{i^*}\right)\right)'\right| + \frac{\ell^3}{d^{\alpha/2}}\sqrt{\frac{8}{\pi}}\frac{K^2}{2} + \frac{3}{4}\frac{\ell^4}{d^{\alpha}}K^2 + \frac{\ell^5}{d^{3\alpha/2}}\sqrt{\frac{32}{\pi}}\frac{K^2}{6} \\ &\quad + \left|\left(\log f\left(X_{i^*}\right)\right)'\right|\left(\frac{\ell^3}{d^{\alpha/2}}\sqrt{\frac{8}{\pi}}\frac{K}{2} + \frac{3}{4}\frac{\ell^4}{d^{\alpha}}K + \frac{\ell^5}{d^{3\alpha/2}}\sqrt{\frac{32}{\pi}}\frac{K}{6}\right) \\ &\quad + \frac{3}{2}\frac{\ell^4}{d^{\alpha}}K + \sqrt{\frac{32}{\pi}}\frac{\ell^5}{d^{3\alpha/2}}\frac{K}{2} + \frac{5}{4}\frac{\ell^6}{d^{2\alpha}}K.\end{aligned}$$

By assumption we have

$$\mathrm{E}\left[\left|\left(\log f\left(X_{i^*}\right)\right)'\right|\right] = \mathrm{E}\left[\left|\frac{f'\left(X_{i^*}\right)}{f\left(X_{i^*}\right)}\right|\right] \leq 1 + \mathrm{E}\left[\left(\frac{f'\left(X_{i^*}\right)}{f\left(X_{i^*}\right)}\right)^4\right] < \infty,$$

so it follows that $\mathrm{E}\left[\left|Gh\left(d, X_{i^*}\right) - \widetilde{G}h\left(d, X_{i^*}\right)\right|\right]$ converges to 0 as $d \to \infty$, which proves the lemma. □

9 Volatility and Drift of the Diffusion

9.1 Convergence to an Approximate Volatility

The aim of the following result is to replace the volatility term of the approximate generator $\widetilde{G}h\left(d, X_{i^*}\right)$ by an asymptotically equivalent, but more convenient expression.

Lemma 8. *We have*

$$\lim_{d\to\infty}\left|\mathrm{E}\left[1\wedge e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)}\right]-\mathrm{E}\left[1\wedge e^{v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)}\right]\right|=0,$$

where $\varepsilon(d,X_j,Y_j)$ *is as in (16) and*

$$\begin{aligned} v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right) &= \sum_{j=1,j\neq i^*}^{n}\varepsilon(d,X_j,Y_j)+\sum_{i=1}^{m}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\frac{d}{dX_j}\log f(\theta_j(d)X_j)(Y_j-X_j)\\ &\quad-\frac{\ell^2}{2d^\alpha}\sum_{i=1}^{m}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\left(\frac{d}{dX_j}\log f(\theta_j(d)X_j)\right)^2. \qquad (20)\end{aligned}$$

Proof. The first step consists in taking the volatility term of $\widetilde{G}h(d,X_{i^*})$ and separating the components whose variance appear only a finite number of times from the other components

$$\begin{aligned}&\mathrm{E}\left[1\wedge e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)}\right]\\ &= \mathrm{E}\left[1\wedge\exp\left\{\sum_{j=1,j\neq i^*}^{n}\varepsilon(d,X_j,Y_j)+\sum_{j=n+1,j\neq i^*}^{d}\left(\log f(\theta_j(d)Y_j)-\log f(\theta_j(d)X_j)\right)\right\}\right].\end{aligned}$$

Writing the difference of log functions as a Taylor expansion with three terms and grouping the components whose scaling term appears infinitely often result in

$$\begin{aligned}&\mathrm{E}\left[1\wedge e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)}\right]\\ &= \mathrm{E}\left[1\wedge\exp\left\{\sum_{j=1,j\neq i^*}^{n}\varepsilon(d,X_j,Y_j)+\sum_{i=1}^{m}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\left[\frac{d}{dX_j}\log f(\theta_j(d)X_j)(Y_j-X_j)\right.\right.\right.\\ &\quad\left.\left.\left.+\frac{1}{2}\frac{d^2}{dX_j^2}\log f(\theta_j(d)X_j)(Y_j-X_j)^2+\frac{1}{6}\frac{d^3}{dU_j^3}\log f(\theta_j(d)U_j)(Y_j-X_j)^3\right]\right\}\right], \qquad (21)\end{aligned}$$

for some $U_i\in(X_i,Y_i)$ or (Y_i,X_i).

We shall now verify if the approximate volatility formed with the function $v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)$ is asymptotically equivalent to the original one. By the triangle's inequality, we have

$$\begin{aligned}&\left|\mathrm{E}\left[1\wedge e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)}\right]-\mathrm{E}\left[1\wedge e^{v\left(d,\mathbf{Y}^{(d-)},\mathbf{X}^{(d-)}\right)}\right]\right|\\ &\leq \mathrm{E}\left[\left|\left(1\wedge e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)}\right)-\left(1\wedge e^{v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)}\right)\right|\right].\end{aligned}$$

By the Lipschitz property of the function $1\wedge e^x$ (see Proposition 2.2 in [14]), and noticing that the first two terms of the function $v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)$ cancel out with the first two terms of the exponential function in (21), we get

$$\mathrm{E}\left[\left|\left(1\wedge e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)}\right)-\left(1\wedge e^{v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)}\right)\right|\right]$$

$$\begin{aligned} &\leq \; \mathrm{E}\left[\left|\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)-v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right|\right] \\ &= \; \mathrm{E}\left[\left|\sum_{i=1}^{m}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\left(\frac{1}{2}\frac{d^2}{dX_j^2}\log f\left(\theta_j\left(d\right)X_j\right)\left(Y_j-X_j\right)^2-\frac{\ell^2}{2d^\alpha}\left(\frac{d}{dX_j}\log f\left(\theta_j\left(d\right)X_j\right)\right)^2\right)\right.\right. \\ &\quad \left.\left.+\frac{1}{6}\sum_{i=1}^{m}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\frac{d^3}{dU_j^3}\log f\left(\theta_j\left(d\right)U_j\right)\left(Y_j-X_j\right)^3\right|\right]. \end{aligned}$$

Since the first double sum forms the random variables $W_i\left(d,\mathbf{X}^{(d)}_{\mathcal{J}(i,d)},\mathbf{Y}^{(d)}_{\mathcal{J}(i,d)}\right)$'s and the derivative appearing in the second term is bounded by some constant, then

$$\begin{aligned} &\mathrm{E}\left[\left|\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)-v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right|\right] \\ &\quad \leq \; \sum_{i=1}^{m}\mathrm{E}\left[\left|W_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)},\mathbf{Y}^{(d)-}_{\mathcal{J}(i,d)}\right)\right|\right]+\sum_{i=1}^{m}c\left(\mathcal{J}\left(i,d\right)\right)\frac{K}{6}\sqrt{\frac{8}{\pi}}\frac{\ell^3}{d^{3\alpha/2}}\frac{d^{3\gamma_i/2}}{K_i^{3/2}}. \end{aligned}$$

Hence by Lemma 6, the RHS converges to 0 as $d\to\infty$, which proves the lemma. □

9.2 Simplified Expression for the Approximate Volatility

Lemma 8 established that on average the acceptance probability of the $(d-1)$-dimensional RWM algorithm is asymptotically equivalent to the function $1\wedge e^{v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)}$. We now wish to find a simpler expression for this new function. This is achieved in the following lemma.

Lemma 9. *If Condition (8) is satisfied, then*

$$\lim_{d\to\infty}\left|\mathrm{E}\left[1\wedge e^{v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)}\right]-2\Phi\left(-\frac{\ell\sqrt{E_R}}{2}\right)\right|=0,$$

where $v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)$ *and* E_R *are as in (20) and (9) respectively.*

Proof. We shall first introduce some notation that will reveal useful for the present proof, as well as for those of the remaining lemmas. For each group of components whose scaling term appears infinitely often in the limit, i.e. for $i=1,\ldots,m$ let

$$R_i\left(d,\mathbf{x}^{(d)-}_{\mathcal{J}(i,d)}\right)=\frac{1}{d^\alpha}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\left(\frac{d}{dx_j}\log\theta_j\left(d\right)f\left(\theta_j\left(d\right)x_j\right)\right)^2. \tag{22}$$

Under this definition, note that the last term of $v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)$ in (20) can be expressed as $-\ell^2\sum_{i=1}^{m}R_i\left(d,\mathbf{x}^{(d)-}_{\mathcal{J}(i,d)}\right)/2$.

Making use of conditioning allows us to express the expectation involved in the limit as

$$\begin{aligned} & \mathrm{E}\left[1 \wedge \exp\left(v\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right)\right)\right] \\ &= \mathrm{E}_{\mathbf{Y}^{(n)-}, \mathbf{X}^{(d)-}}\left[\mathrm{E}_{\mathbf{Y}^{(d-n)-}}\left[1 \wedge \exp\left(v\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right)\right)\right]\right]. \end{aligned} \tag{23}$$

To solve the inner expectation, we need to find the distribution of $v\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right) \left| \mathbf{Y}^{(n)-}, \mathbf{X}^{(d)-}\right.$. Since $(Y_j - X_j) \,|X_j\,, j = 1, \cdots, d$, are *iid* and normally distributed with mean 0 and variance ℓ^2/d^α, then

$$\begin{aligned} & v\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right) \left| \mathbf{Y}^{(n)-}, \mathbf{X}^{(d)-}\right. \\ &\sim N\left(\sum_{j=1, j\neq i^*}^{n} \varepsilon\left(d, X_j, Y_j\right) - \frac{\ell^2}{2}\sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right),\ \ell^2 \sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\right). \end{aligned}$$

Applying Proposition 2.4 in [14] allows to express the inner expectation in (23) in terms of $\Phi(\cdot)$, the *cdf* of a standard normal random variable

$$\begin{aligned} & \mathrm{E}_{\mathbf{Y}^{(d-n)-}}\left[1 \wedge e^{v\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right)}\right] \\ &= \Phi\left(\frac{\sum_{j=1, j\neq i^*}^{n} \varepsilon\left(d, X_j, Y_j\right) - \frac{\ell^2}{2}\sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}{\sqrt{\ell^2 \sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}}\right) \\ &\quad + \exp\left(\sum_{j=1, j\neq i^*}^{n} \varepsilon\left(d, X_j, Y_j\right)\right) \Phi\left(\frac{-\sum_{j=1, j\neq i^*}^{n} \varepsilon\left(d, X_j, Y_j\right) - \frac{\ell^2}{2}\sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}{\sqrt{\ell^2 \sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}}\right) \\ &\equiv M\left(d, \mathbf{Y}^{(n)-}, \mathbf{X}^{(d)-}\right). \end{aligned}$$

We are then left to evaluate the expectation of $M(\cdot)$. Again using conditional expectations

$$\mathrm{E}\left[M\left(d, \mathbf{Y}^{(n)-}, \mathbf{X}^{(d)-}\right)\right] = \mathrm{E}_{\mathbf{X}^{(d-n)-}}\left[\mathrm{E}_{\mathbf{Y}^{(n)-}, \mathbf{X}^{(n)-}}\left[M\left(d, \mathbf{Y}^{(n)-}, \mathbf{X}^{(d)-}\right)\right]\right],$$

From Proposition 12, we find that both terms forming the function $M\left(d, \mathbf{Y}^{(n)-}, \mathbf{X}^{(d)-}\right)$ have the same inner expectation. The unconditional expectation thus simplifies to

$$\mathrm{E}\left[M\left(d, \mathbf{Y}^{(n)-}, \mathbf{X}^{(d)-}\right)\right] = 2\mathrm{E}\left[\Phi\left(\frac{\sum_{j=1, j\neq i^*}^{n} \varepsilon\left(d, X_j, Y_j\right) - \frac{\ell^2}{2}\sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}{\sqrt{\ell^2 \sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}}\right)\right].$$

We now find the limit of the term inside the function $\Phi(\cdot)$ as $d \to \infty$. From Proposition 13, $\varepsilon(d, X_j, Y_j)$ converges in probability to 0 for all $j \in \{1, \ldots, n\}$ but excluding $j = i^*$. Similarly, we use Proposition 14 to conclude that $\sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right) \to_p E_R$. Furthermore, $E_R > 0$ since there is at least one $i \in \{1, \ldots, m\}$ such that $\lim_{d\to\infty} c\left(\mathcal{J}(i,d)\right) d^{\gamma_i}/d^\alpha > 0$.

By applying Slutsky's Theorem, the Continuous Mapping Theorem and by recalling that convergence in probability and convergence in distribution are equivalent when the limit is a constant, we conclude that

$$\Phi\left(\frac{\sum_{j=1,j\neq i^*}^{n}\varepsilon\left(d,X_j,Y_j\right)-\frac{\ell^2}{2}\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)}\right)}{\sqrt{\ell^2\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)}\right)}}\right)\rightarrow_p \Phi\left(-\frac{\ell\sqrt{E_R}}{2}\right).$$

Since $\Phi(\cdot)$ is positive and bounded by 1, we finally use the Bounded Convergence Theorem to find

$$\begin{aligned}\mathrm{E}\left[1\wedge e^{v\left(d,\mathbf{Y}^{(d-)},\mathbf{X}^{(d-)}\right)}\right] &= 2\mathrm{E}\left[\Phi\left(\frac{\sum_{j=1,j\neq i^*}^{n}\varepsilon\left(d,X_j,Y_j\right)-\frac{\ell^2}{2}\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)}\right)}{\sqrt{\ell^2\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)}\right)}}\right)\right] \\ &\rightarrow 2\Phi\left(-\frac{\ell\sqrt{E_R}}{2}\right) \quad \text{as } d\rightarrow\infty,\end{aligned}$$

which completes the proof of the lemma. □

9.3 Convergence to an Approximate Drift

The following result aims to replace the drift term of the approximate generator $\widetilde{G}h\left(d,X_{i^*}\right)$ in Lemma 7 by an asymptotically equivalent, but more convenient expression.

Lemma 10. *We have*

$$\begin{aligned}\lim_{d\rightarrow\infty}\Bigg|\mathrm{E}\left[e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)};\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)<0\right] \\ -\mathrm{E}\left[e^{v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)};v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)<0\right]\Bigg|=0,\end{aligned} \qquad (24)$$

where the functions $\varepsilon\left(d,X_j,Y_j\right)$ and $v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)$ are as in (16) and (20) respectively.

Proof. First, let

$$T(x) = e^x\mathbf{1}_{(x<0)}=\begin{cases}e^x, & x<0\\ 0, & x\geq 0\end{cases}.$$

It is important to realize that the function $T(x)$ is not Lipschitz, which keeps us from reproducing the proof of Lemma 8. The approach we use is to show that

$$T\left(\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)\right)\rightarrow_p T\left(v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right), \qquad (25)$$

and then use this result to prove convergence of expectations.

Let

$$A\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right) = T\left(\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)\right)-T\left(v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right)$$

and

$$\delta\left(d\right) = \left(\sum_{i=1}^{m}\mathrm{E}\left[\left|W_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)},\mathbf{Y}^{(d)-}_{\mathcal{J}(i,d)}\right)\right|\right]+\sum_{i=1}^{m}c\left(\mathcal{J}\left(i,d\right)\right)\frac{K}{6}\sqrt{\frac{8}{\pi}}\frac{\ell^3}{d^{3\alpha/2}}\frac{d^{3\gamma_i/2}}{K_i^{3/2}}\right)^{1/2}.$$

In order to simplify the expressions involved in the following development, we shall omit the arguments $\mathbf{Y}^{(d)-}$ and $\mathbf{X}^{(d)-}$ in the functions $\varepsilon(\cdot)$, $v(\cdot)$ and $A(\cdot)$. We have

$$\begin{aligned}&\mathrm{P}\left(\left|A\left(d\right)\right|\geq\delta\left(d\right)\right)\\ &= \mathrm{P}\left(\left|A\left(d\right)\right|\geq\delta\left(d\right);\sum\varepsilon\left(d\right)\geq0;v\left(d\right)\geq0\right)+\mathrm{P}\left(\left|A\left(d\right)\right|\geq\delta\left(d\right);\sum\varepsilon\left(d\right)<0;v\left(d\right)<0\right)\\ &\quad+\mathrm{P}\left(\left|A\left(d\right)\right|\geq\delta\left(d\right);\sum\varepsilon\left(d\right)\geq0;v\left(d\right)<0\right)+\mathrm{P}\left(\left|A\left(d\right)\right|\geq\delta\left(d\right);\sum\varepsilon\left(d\right)<0;v\left(d\right)\geq0\right).\end{aligned}$$

We can bound the third term on the RHS by

$$\begin{aligned}&\mathrm{P}\left(\left|A\left(d\right)\right|\geq\delta\left(d\right);\sum\varepsilon\left(d\right)\geq0;v\left(d\right)<0\right)\\ &\leq \mathrm{P}\left(\sum\varepsilon\left(d\right)\geq0;v\left(d\right)<0;\left|\sum\varepsilon\left(d\right)-v\left(d\right)\right|<\delta\left(d\right)\right)\\ &\quad+\mathrm{P}\left(\sum\varepsilon\left(d\right)\geq0;v\left(d\right)<0;\left|\sum\varepsilon\left(d\right)-v\left(d\right)\right|\geq\delta\left(d\right)\right)\end{aligned}$$

and similarly for the fourth term. Also note that if $\sum\varepsilon(d)\geq0$ and $v(d)\geq0$, or $\sum\varepsilon(d)<0$ and $v(d)<0$, then

$$\left|A\left(d\right)\right|\leq\left|\sum\varepsilon\left(d\right)-v\left(d\right)\right|.$$

Therefore,

$$\begin{aligned}\mathrm{P}\left(\left|A\left(d\right)\right|\geq\delta\left(d\right)\right) &\leq \mathrm{P}\left(\left|\sum\varepsilon\left(d\right)-v\left(d\right)\right|\geq\delta\left(d\right)\right)\\ &\quad+\mathrm{P}\left(\sum\varepsilon\left(d\right)\geq0;v\left(d\right)<0;\left|\sum\varepsilon\left(d\right)-v\left(d\right)\right|<\delta\left(d\right)\right)\\ &\quad+\mathrm{P}\left(\sum\varepsilon\left(d\right)<0;v\left(d\right)\geq0;\left|\sum\varepsilon\left(d\right)-v\left(d\right)\right|<\delta\left(d\right)\right).\end{aligned}$$

Since $\sum\varepsilon(d)$ and $v(d)$ are of different sign but the difference between them must be less than $\delta(d)$, we can bound the last two terms and obtain

$$\begin{aligned}\mathrm{P}\left(\left|A\left(d\right)\right|\geq\delta\left(d\right)\right) &\leq \mathrm{P}\left(\left|\sum\varepsilon\left(d\right)-v\left(d\right)\right|\geq\delta\left(d\right)\right)\\ &\quad+\mathrm{P}\left(-\delta\left(d\right)<v\left(d\right)<0\right)+\mathrm{P}\left(0\leq v\left(d\right)<\delta\left(d\right)\right)\\ &= \mathrm{P}\left(\left|\sum\varepsilon\left(d\right)-v\left(d\right)\right|\geq\delta\left(d\right)\right)+\mathrm{P}\left(-\delta\left(d\right)<v\left(d\right)<\delta\left(d\right)\right). \qquad (26)\end{aligned}$$

By Markov's inequality and the proof of Lemma 8 the first term on the RHS of (26) satisfies

$$\begin{aligned} &\mathrm{P}\left(\left|\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)-v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right|\geq\delta\left(d\right)\right)\\ &\quad\leq\frac{1}{\delta\left(d\right)}\mathrm{E}\left[\left|\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)-v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right|\right]\leq\sqrt{\delta\left(d\right)}\to 0\ \text{ as } d\to\infty. \end{aligned}$$

Now consider the second term on the RHS of (26). From the proof of Lemma 9, we know the distribution of $v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\left|\mathbf{Y}^{(n)-},\mathbf{X}^{(d)-}\right.$. Using conditional theory, we have

$$\mathrm{P}\left(\left|v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right|<\delta\left(d\right)\right)\ =\ \mathrm{E}_{\mathbf{Y}^{(n)-},\mathbf{X}^{(d)-}}\left[\mathrm{P}_{\mathbf{Y}^{(d-n)-}}\left(\left|v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right|<\delta\left(d\right)\right)\right].$$

Focusing on the conditional probability, we write

$$\begin{aligned} &\mathrm{P}_{\mathbf{Y}^{(d-n)-}}\left(\left|v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right|<\delta\left(d\right)\right)\\ &=\ \Phi\left(\frac{\delta\left(d\right)-\sum_{j=1,j\neq i^*}^{n}\varepsilon\left(d,X_j,Y_j\right)+\frac{\ell^2}{2}\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)-}\right)}{\ell\sqrt{\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)-}\right)}}\right)\\ &\quad-\Phi\left(\frac{-\delta\left(d\right)-\sum_{j=1,j\neq i^*}^{n}\varepsilon\left(d,X_j,Y_j\right)+\frac{\ell^2}{2}\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)-}\right)}{\ell\sqrt{\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)-}\right)}}\right). \end{aligned}$$

Using the convergence results developed in the proof of Lemma 9 along with the fact that $\delta\left(d\right)\to 0$ as $d\to\infty$, we deduce that $\mathrm{P}_{\mathbf{Y}^{(d-n)-}}\left(\left|v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right|<\delta\left(d\right)\right)\to_p 0$. Using the Bounded Convergence Theorem, we then find that the unconditional probability $\mathrm{P}\left(\left|v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right|<\delta\left(d\right)\right)\to_p 0$ as well. Since we showed that $P\left(\left|A\left(d\right)\right|\geq\delta\left(d\right)\right)\to 0$ as $d\to\infty$, this means that (25) is true and thus (24) can be verified with the Bounded Convergence Theorem. □

9.4 Simplified Expression for the Approximate Drift

The goal of this section is to determine a simpler expression for the approximate drift term introduced in Lemma 7.

Lemma 11. *If Condition (8) is satisfied, then*

$$\lim_{d\to\infty}\left|\mathrm{E}\left[e^{v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)};v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)<0\right]-\Phi\left(-\frac{\ell\sqrt{E_R}}{2}\right)\right|=0,$$

where the functions $\varepsilon\left(d,X_j,Y_j\right)$ *and* $v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)$ *are as in (16) and (20) respectively.*

Proof. The proof is similar to that of Lemma 9 and for this reason, we just outline the differences. We know the distribution of $v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\left|\mathbf{Y}^{(n)-},\mathbf{X}^{(d)-}\right.$ from the proof of Lemma 9, so we can use Proposition 2.4 in [14] to obtain

$$\begin{aligned}
&\mathrm{E}_{\mathbf{Y}^{(d-n)-}}\left[e^{v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)};v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)<0\right]\\
&= \exp\left(\sum_{j=1,j\neq i^*}^{n}\varepsilon\left(d,X_j,Y_j\right)\right)\Phi\left(\frac{-\sum_{j=1,j\neq i^*}^{n}\varepsilon\left(d,X_j,Y_j\right)-\frac{\ell^2}{2}\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)-}\right)}{\sqrt{\ell^2\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)-}\right)}}\right).
\end{aligned}$$

Applying Proposition 12, we find

$$\begin{aligned}
&\mathrm{E}\left[e^{v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)};v\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)<0\right]\\
&= \mathrm{E}\left[\Phi\left(\frac{\sum_{j=1,j\neq i^*}^{n}\varepsilon\left(d,X_j,Y_j\right)-\frac{\ell^2}{2}\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)-}\right)}{\sqrt{\ell^2\sum_{i=1}^{m}R_i\left(d,\mathbf{X}_{\mathcal{J}(i,d)}^{(d)-}\right)}}\right)\right].
\end{aligned}$$

Repeating the reasoning of the proof of Lemma 9 completes the demonstration of the present lemma. □

10 Discussion

The theorems in this paper basically extend the *iid* work of [14] to a more general setting where the scaling term of each target component is allowed to depend on the dimension of the target distribution. The conclusions achieved in these theorems are very similar to those in [14] in the sense that the obtained asymptotically optimal acceptance rates are identical, the only difference lying in the optimal scaling values themselves. However, crucial to the validity of these results is the fulfilment of an important condition; condition (8) is in fact the key point ensuring that the process will asymptotically behave as in the *iid* case. The intuition behind this statement is that there will be no component converging significantly faster than the others, justifying the regular asymptotic behavior of the algorithm. This work thus partially answers Open Problem #3 of [17].

The well-known acceptance rate 0.234 has long be believed to hold under certain perturbations of the target density. The particularity of our results is that they provide, for the specified target setting, a necessary and sufficient condition under which the optimality of 0.234 is verified. Moreover, they allow determining with certitude whether or not this acceptance rate is optimal for virtually any correlated multivariate normal target distribution. Contrarily to what seemed to be a common belief, multivariate normal distributions do not always adopt a conventional limiting behavior. There indeed exist cases where the asymptotically optimal acceptance rate is significantly smaller than 0.234, which is discussed in [1].

It was shown in the *iid* case that even though the results are of asymptotic nature, they are also pretty accurate in small dimensions ($d \geq 10$). In the present case however, this fact is not always verified and care must be exercised in practice. In particular, if there exists a finite number of scaling terms such that λ_j is close to α (but with $\lambda_j < \alpha$ of course, otherwise Condition (8) would be violated) then the optimal acceptance rate converges extremely slowly to 0.234 from above. For instance, suppose that the variances of a d-dimensional multivariate normal target with independent components are $\left(d^{-\lambda}, 1, \ldots, 1\right)$, where $\lambda < 1$. The proposal scaling is then of the form $\sigma^2(d) = \ell^2/d$ and the closer to 1 is λ, the slower is the convergence of the optimal acceptance rate to 0.234. In fact, for $\lambda = 0.75$, simulations show that d must be as big as $200,000$ for the optimal acceptance rate to be reasonably close to 0.234. Simulations also show that for $\alpha - \lambda \leq 0.5$, the asymptotic results are accurate in relatively small dimensions, just as in the *iid* case. Detailed examples and simulation studies illustrating this paper's results as well as those introduced in [1] are presented in [2].

Appendix

The following results are useful for proving lemmas of Section 9. The first result demonstrates the equivalence between two expectations. The second and third propositions aim to prove the convergence in probability of some variables to a constant.

Proposition 12. *Let $\mathbf{X}_j$ be distributed according to the density $\theta_j(d) f(\theta_j(d) x_j)$ for $j = 1, \ldots, d$. Also let $\mathbf{Y}^{(d)} \big| \mathbf{X}^{(d)} \sim N\left(\mathbf{X}^{(d)}, \sigma^2(d) I_{d\times d}\right)$ and $\varepsilon(d, X_j, Y_j)$ as in (16). We have*

$$\mathrm{E}_{\mathbf{Y}^{(n)-},\mathbf{X}^{(n)-}}\left[\exp\left(\sum_{j=1,j\neq i^*}^{n}\varepsilon(d,X_j,Y_j)\right)\Phi\left(\frac{-\sum_{j=1,j\neq i^*}^{n}\varepsilon(d,X_j,Y_j)-\frac{\ell^2}{2}\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}{\sqrt{\ell^2\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}}\right)\right]$$
$$= \mathrm{E}_{\mathbf{Y}^{(n)-},\mathbf{X}^{(n)-}}\left[\Phi\left(\frac{\sum_{j=1,j\neq i^*}^{n}\varepsilon(d,X_j,Y_j)-\frac{\ell^2}{2}\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}{\sqrt{\ell^2\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}}\right)\right].$$

Proof. Developing the first expectation and simplifying the integrand yield

$$\mathrm{E}_{\mathbf{Y}^{(n)-},\mathbf{X}^{(n)-}}\left[\prod_{j=1,j\neq i^*}^{n}\frac{f(\theta_j(d)Y_j)}{f(\theta_j(d)X_j)}\Phi\left(\frac{-\log\prod_{j=1,j\neq i^*}^{n}\frac{f(\theta_j(d)Y_j)}{f(\theta_j(d)X_j)}-\frac{\ell^2}{2}\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}{\sqrt{\ell^2\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}}\right)\right]$$
$$= \int\int\Phi\left(\frac{\log\prod_{j=1,j\neq i^*}^{n}\frac{f(\theta_j(d)x_j)}{f(\theta_j(d)y_j)}-\frac{\ell^2}{2}\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}{\sqrt{\ell^2\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)}}\right)$$
$$\prod_{j=1,j\neq i^*}^{n}\theta_j(d)f(\theta_j(d)y_j)\,C\exp\left(-\frac{1}{2\sigma^2(d)}\sum_{j=1,j\neq i^*}^{n}(x_j-y_j)^2\right)d\mathbf{y}^{(n)-}d\mathbf{x}^{(n)-}.$$

Since the integrand is positive we can use Fubini's Theorem to change the order of integration. Substituting $\mathbf{y}^{(n-)}$ for $\mathbf{x}^{(n-)}$ and vice versa then yields the desired result. □

Proposition 13. *Let*

$$\varepsilon(d, X_j, Y_j) = \log \frac{f(\theta_j(d) Y_j)}{f(\theta_j(d) X_j)},$$

where $\theta_j(d) = K_j/d^{\lambda_j}$ *for* $j \in \{1, \ldots, n\}$. *If* $\lambda_j < \alpha$, *then* $\varepsilon(d, X_j, Y_j) \to_p 0$.

Proof. By Taylor's Theorem, we have the following three-term expansion

$$\begin{aligned}\varepsilon(d, X_j, Y_j) &= (\log f(\theta_j(d) X_j))'(Y_j - X_j) + \frac{1}{2}(\log f(\theta_j(d) X_j))''(Y_j - X_j)^2 \\ &\quad + \frac{1}{6}(\log f(\theta_j(d) U_j))'''(Y_j - X_j)^3,\end{aligned}$$

for some $U_j \in (X_j, Y_j)$ or (Y_j, X_j).

Using conditional expectations, changes of variables and the triangle's inequality, we find

$$\mathrm{E}[\varepsilon(d, X_j, Y_j)] \leq \frac{\ell^2}{2d^\alpha}\theta_j^2(d)\,\mathrm{E}\left[\left|(\log f(X))''\right|\right] + \frac{1}{6}\theta_j^3(d)\,\mathrm{E}\left[\left|(\log f(U))'''\right| |Y_j - X_j|^3\right].$$

Since $\left|(\log f(X))''\right|$ and $\left|(\log f(U))'''\right|$ are bounded by some constant (say K) and since $\lambda_j < \alpha$, we have

$$\begin{aligned}\mathrm{E}[\varepsilon(d, X_j, Y_j)] &\leq \frac{\ell^2}{2d^\alpha}\frac{d^{\lambda_j}}{K_j}K + \frac{1}{6}\sqrt{\frac{8}{\pi}}\frac{\ell^3}{d^{3\alpha/2}}\left(\frac{d^{\lambda_j}}{K_j}\right)^{3/2}K \\ &\to 0 \quad \text{as } d \to \infty,\end{aligned}$$

We now use the previous Taylor expansion to bound the variance

$$\begin{aligned}&\mathrm{Var}(\varepsilon(d, X_j, Y_j)) \\ &\quad\leq \mathrm{E}\left[\varepsilon^2(d, X_j, Y_j)\right] \\ &\quad= \mathrm{E}\Bigg[\Bigg((\log f(\theta_j(d) X_j))'(Y_j - X_j) + \frac{1}{2}(\log f(\theta_j(d) X_j))''(Y_j - X_j)^2 \\ &\qquad + \frac{1}{6}(\log f(\theta_j(d) U_j))'''(Y_j - X_j)^3\Bigg)^2\Bigg].\end{aligned}$$

Developing the square, applying changes of variables and using conditional expectations result in

$$\begin{aligned}\mathrm{E}\left[\varepsilon^2(d, X_j, Y_j)\right] &= \frac{\ell^2}{d^\alpha}\theta_j^2(d)\,\mathrm{E}\left[\left((\log f(X))'\right)^2\right] + \frac{3}{4}\frac{\ell^4}{d^{2\alpha}}\theta_j^4(d)\,\mathrm{E}\left[\left[(\log f(X))''\right]^2\right] \\ &\quad + \frac{1}{36}\theta_j^6(d)\,\mathrm{E}\left[\left[(\log f(U))'''\right]^2 (Y_j - X_j)^6\right]\end{aligned}$$

$$+\sqrt{\frac{8}{\pi}}\frac{\ell^3}{d^{3\alpha/2}}\theta_j^3(d)\,\mathrm{E}\left[(\log f(X))'(\log f(X))''\right]$$
$$+\frac{1}{3}\theta_j^4(d)\,\mathrm{E}_X\left[(\log f(X))'\,\mathrm{E}_{Y_j}\left[(\log f(U))'''(Y_j-X_j)^4\right]\right]$$
$$+\frac{1}{6}\theta_j^5(d)\,\mathrm{E}_X\left[(\log f(X))''\,\mathrm{E}_{Y_j}\left[(\log f(U))'''(Y_j-X_j)^5\right]\right].$$

By the triangle's inequality, and again using the fact that $\left|(\log f(X))''\right|$ and $\left|(\log f(U))'''\right|$ are bounded, we obtain

$$\begin{aligned}\mathrm{E}\left[\varepsilon^2(d,X_j,Y_j)\right] \;&\leq\; \frac{\ell^2}{d^\alpha}\frac{d^{\lambda_j}}{K_j}\mathrm{E}\left[\left((\log f(X))'\right)^2\right]+\frac{3}{4}\frac{\ell^4}{d^{2\alpha}}\frac{d^{2\lambda_j}}{K_j^2}K^2+\frac{15}{36}\frac{d^{3\lambda_j}}{K_j^3}\frac{\ell^6}{d^{3\alpha}}K^2\\ &+\sqrt{\frac{8}{\pi}}\frac{\ell^3}{d^{3\alpha/2}}\frac{d^{3\lambda_j/2}}{K_j^{3/2}}K\mathrm{E}\left[\left|(\log f(X))'\right|\right]\\ &+\frac{d^{2\lambda_j}}{K_j^2}\frac{\ell^4}{d^{2\alpha}}K\mathrm{E}\left[\left|(\log f(X))'\right|\right]+\frac{5}{2}\frac{d^{5\lambda_j/2}}{K_j^{5/2}}\frac{\ell^5}{d^{5\alpha/2}}K^2.\end{aligned}$$

By assumption, $\mathrm{E}\left[\left((\log f(X))'\right)^2\right]$ is bounded by some finite constant. Since $\lambda_j<\alpha$, the previous expression converges to 0 as $d\to\infty$.

To complete the proof of the proposition we use Chebychev's inequality and find that for all $\epsilon>0$

$$\mathrm{P}\left(|\varepsilon(d,X_j,Y_j)|\geq\epsilon\right)\;\leq\;\frac{1}{\epsilon^2}\mathrm{Var}\left(\varepsilon(d,X_j,Y_j)\right)\leq\frac{1}{\epsilon^2}\mathrm{E}\left[\varepsilon^2(d,X_j,Y_j)\right]\to 0\;\text{ as } d\to\infty.$$

□

Proposition 14. *Let $R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)$ be as in (22), with $i\in\{1,\ldots,m\}$. We then have $\sum_{i=1}^m R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\to_p E_R$, where E_R is as in (9).*

Proof. The expectation of each variable satisfies

$$\begin{aligned}&\mathrm{E}\left[R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\right]\\ &=\frac{1}{d^\alpha}\int_{\mathbf{R}}\cdots\int_{\mathbf{R}}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\left(\frac{d}{dx_j}\log\theta_j(d)f(\theta_j(d)x_j)\right)^2\prod_{k\in\mathcal{J}(i,d),k\neq i^*}\theta_k(d)f(\theta_k(d)x_k)\,dx_k\\ &=\frac{\theta_{n+i}^2(d)}{d^\alpha}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\int_{\mathbf{R}}\left(\frac{f'(\theta_{n+i}(d)x_j)}{f(\theta_{n+i}(d)x_j)}\right)^2\theta_j(d)f(\theta_j(d)x_j)\,dx_j\\ &=\frac{d^{\gamma_i}}{K_{n+i}d^\alpha}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\int_{\mathbf{R}}\left(\frac{f'(x)}{f(x)}\right)^2 f(x)\,dx,\end{aligned}$$

and writing the integral as an expectation yields

$$\mathrm{E}\left[R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\right] = \frac{c\left(\mathcal{J}\left(i,d\right)\right)}{d^{\alpha}}\frac{d^{\gamma_i}}{K_{n+i}}\mathrm{E}\left[\left(\frac{f'\left(X\right)}{f\left(X\right)}\right)^2\right]. \tag{27}$$

In the limit, the expectation of the sum of variables then becomes

$$E_R = \lim_{d\to\infty}\sum_{i=1}^{m}\mathrm{E}\left[R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\right] = \lim_{d\to\infty}\sum_{i=1}^{m}\frac{c\left(\mathcal{J}\left(i,d\right)\right)}{d^{\alpha}}\frac{d^{\gamma_i}}{K_{n+i}}\mathrm{E}\left[\left(\frac{f'\left(X\right)}{f\left(X\right)}\right)^2\right],$$

which in the present case is positive but finite.

Since all X_j's are independent, the variance of this sum is given by

$$\mathrm{Var}\left(\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\right) = \sum_{i=1}^{m}\frac{1}{d^{2\alpha}}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\mathrm{Var}\left(\left[\left(\log\theta_j\left(d\right)f\left(\theta_j\left(d\right)X_j\right)\right)'\right]^2\right),$$

and using the fact that $\mathrm{Var}\left(X\right)\leq\mathrm{E}\left[X^2\right]$ along with a change of variable yield

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\right) &\leq \sum_{i=1}^{m}\frac{1}{d^{2\alpha}}\sum_{j\in\mathcal{J}(i,d),j\neq i^*}\theta_j^4\left(d\right)\mathrm{E}\left[\left[\left(\log f\left(X\right)\right)'\right]^4\right]\\ &= \sum_{i=1}^{m}\frac{1}{d^{2\alpha}}\frac{d^{2\gamma_i}}{K_{n+i}^2}c\left(\mathcal{J}\left(i,d\right)\right)\mathrm{E}\left[\left(\frac{f'\left(X\right)}{f\left(X\right)}\right)^4\right].\end{aligned}$$

By assumption, we know that the expectation involved in the previous expression is finite. Since $c\left(\mathcal{J}\left(i,d\right)\right)d^{\gamma_i}\leq d^{\alpha}$ and $c\left(\mathcal{J}\left(i,d\right)\right)\to\infty$ as $d\to\infty$ for $i=1,\ldots,m$, the variance thus converges to 0 as $d\to\infty$.

To conclude the proof of the lemma, we use Chebychev's inequality and find that for all $\epsilon>0$

$$\mathrm{P}\left(\left|\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)-E_R\right|\geq\epsilon\right) \leq \frac{1}{\epsilon^2}\mathrm{Var}\left(\sum_{i=1}^{m}R_i\left(d,\mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\right)\to 0 \quad \text{as} \quad d\to\infty.$$

□

Acknowledgments

This work is part of my Ph.D. thesis and has been supported by NSERC of Canada. Special thanks are due to my supervisor, Professor Jeffrey S. Rosenthal, without who this work would not have been complete. His expertise, guidance and encouragements have been precious throughout my studies. I also acknowledge useful conversations with Professor Gareth O. Roberts.

References

[1] Bédard, M. (2006). Optimal Acceptance Rates for Metropolis Algorithms: Moving Beyond 0.234. Submitted for publication.

[2] Bédard, M. (2006). Efficient Sampling using Metropolis Algorithms: Applications of Optimal Scaling Results. Submitted for publication.

[3] Besag, J., Green, P.J. (1993). Spatial statistics and Bayesian computation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **55**, 25-38.

[4] Besag, J., Green, P.J., Higdon, D., Mergensen, K. (1995). Bayesian computation adn stochastic systems. *Statist. Sci.* **10**, 3-66.

[5] Billingsley, P. (1995). *Probability and Measure, 3rd ed.* John Wiley & Sons, New York.

[6] Breyer, L.A., Piccioni, M., Scarlatti, S. (2002). Optimal Scaling of MALA for Nonlinear Regression. *Ann. Appl. Probab.* **14**, 1479-1505.

[7] Breyer, L.A., Roberts, G.O. (2000). From Metropolis to Diffusions: Gibbs States and Optimal Scaling. *Stochastic Process. Appl.* **90**, 181-206.

[8] Christensen, O.F., Roberts, G.O., Rosenthal, J.S. (2003). Scaling Limits for the Transient Phase of Local Metropolis-Hastings Algorithms. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 253-69.

[9] Ethier, S.N., Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence.* Wiley.

[10] Grimmett, G.R., Stirzaker, D.R. (1992). *Probability and Random Processes.* Oxford.

[11] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika.* **57**, 97-109.

[12] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087-92.

[13] Neal, P., Roberts, G.O. (2004). Optimal Scaling for Partially Updating MCMC Algorithms. *To appear in Ann. Appl. Probab.*

[14] Roberts, G.O., Gelman, A., Gilks, W.R. (1997). Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *Ann. Appl. Probab.* **7**, 110-20.

[15] Roberts, G.O., Rosenthal, J.S. (1998). Optimal Scaling of Discrete Approximations to Langevin Diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60**, 255-68.

[16] Roberts, G.O., Rosenthal, J.S. (2001). Optimal Scaling for various Metropolis-Hastings algorithms. *Statist. Sci.* **16**, 351-67.

[17] Roberts, G.O., Rosenthal, J.S. (2004). General State Space Markov Chains and MCMC Algorithms. *Probab. Surveys* **1**, 20-71.

[18] Rosenthal, J.S. (2000). *A First Look at Rigorous Probability Theory.* World Scientific, Singapore.