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

Mylène Bédard *

Abstract

In this paper, we shall optimize the efficiency of 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 with a scaling term significantly smaller than the others, the asymptotically optimal acceptance rate is the well-known 0.234.

Keywords: Metropolis algorithm, weak convergence, optimal scaling, diffusion, Markov chain Monte Carlo

1 Introduction

Metropolis algorithms ([9], [8]) provide a method to sample from highly complex probability distributions. The ease of implementation and wide applicability of these algorithms have conferred them their popularity, and they are frequently used nowadays by all levels of practitioners in various fields of application. However, their convergence can sometimes be lengthy, which calls for an optimization of their performance. Because the efficiency of Metropolis algorithms depends crucially on the scaling of the proposal density chosen for their implementation, it is thus fundamental to judiciously choose this parameter.

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 [11]. In particular, the authors considered d-dimensional target distributions with *iid* components and studied the asymptotic behavior (as $d \to \infty$) of Metropolis 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 (AOAR) makes these theoretical results extremely useful in practice. Optimal scaling issues have been explored by other authors, namely [12], [6], [13], [5] and [10].

*Department of Statistics, University of Warwick, Coventry, U.K., CV4 7AL.
Email: M.bedard@warwick.ac.uk Submitted January 2006, accepted December 2006.
AMS Class: primary, 60F05 ; secondary, 65C40

In this paper, we carry out a similar study for d-dimensional target distributions with independent components. The particularity of our model is that the scaling term of each component is allowed to depend on the dimension of the target distribution, which constitutes a critical distinction with the *iid* case. We provide a condition under which the algorithm admits the same limiting diffusion process and the same AOAR as those found in [11]. This is achieved in the first place by determining the appropriate form for the proposal scaling as a function of d, 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 (appropriately rescaled) converges to a Langevin diffusion process with a certain speed measure. Obtaining the AOAR is thus a simple matter of optimizing the speed measure of the diffusion.

The paper is structured as follows. In Section 2, we describe the Metropolis algorithm and introduce the target distribution setting. The main results are presented in Section 3, along with a discussion about inhomogeneous proposal distributions and some extensions. We prove the theorems in Section 4 using lemmas proved in Sections 5 and 6, finally concluding the paper with a discussion.

2 Sampling from the Target Distribution

2.1 The Metropolis Algorithm

The idea behind the Metropolis algorithm is 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 Lebesgue measure. Also let the proposed moves be normally distributed around $\mathbf{x}$, i.e. $N\left(\mathbf{x}, \sigma^2 I_d\right)$ for some σ^2 and with I_d the d-dimensional identity matrix. The Metropolis 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 the normal density $q\left(\mathbf{X}_t, \mathbf{y}\right) d\mathbf{y}$. The probability of accepting the proposed value $\mathbf{Y}_{t+1}$ as the new value for the chain is $\alpha\left(\mathbf{X}_t, \mathbf{Y}_{t+1}\right)$, where

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

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 $\mathbf{X}_{t+1} = \mathbf{X}_t$.

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 in spite of a very high acceptance rate, simulation will move very slowly to the target distribution. 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.

Before finding an appropriate value for σ^2 between these extremes, we first define a criterion which is closely related to the algorithm efficiency. The notion of π-average acceptance rate is defined in [11] as $\mathrm{E}\left[1 \wedge \frac{\pi(\mathbf{Y})}{\pi(\mathbf{X})}\right] = \int\int \pi\left(\mathbf{x}\right) \alpha\left(\mathbf{x}, \mathbf{y}\right) q\left(\mathbf{x}, \mathbf{y}\right) d\mathbf{x}d\mathbf{y}$ for the d-dimensional

Metropolis algorithm.

2.2 The Target Distribution

Consider the following d-dimensional target density

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

We impose the following regularity conditions on the density f: f is a positive C^2 function and $(\log f(x))'$ is Lipschitz continuous. We also suppose that $\mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^4\right] = \int_{\mathbf{R}} \left(\frac{f'(x)}{f(x)}\right)^4 f(x)\, dx < \infty$ and $\mathrm{E}\left[\left(\frac{f''(X)}{f(X)}\right)^2\right] < \infty$.

The d target components, although independent, are however not identically distributed. In particular, we consider the case where the scaling terms $\theta_j^{-2}(d)$, $j = 1, \ldots, d$ take the following form

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

Ultimately, we shall be interested in the limit of the target distribution as $d \to \infty$. Let $n < \infty$ denote the number of components whose scaling term appears a finite number of times in the limit of $\mathbf{\Theta}^{-2}(d)$. Also, let the j-th of these n scaling terms be K_j/d^{λ_j}, $j = 1, \ldots, n$, where $\lambda_j \in (-\infty, \infty)$ and 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}, $i = 1, \ldots, m$, with $\gamma_i \in (-\infty, \infty)$. For now, we assume the constants $0 < K_{n+i} < \infty$ to be the same for all scaling terms within each of the m groups. We shall relax this assumption in Section 3.2.

For $i = 1, \ldots, m$, define the sets $\mathcal{J}(i, d) = \left\{j \in \{1, \ldots, d\}; \theta_j^{-2}(d) = \frac{K_{n+i}}{d^{\gamma_i}}\right\}$. The i-th set thus contains positions of components with a scaling term equal to K_{n+i}/d^{γ_i}. These sets are such that $\dot{\bigcup}_{i=1}^{m} \mathcal{J}(i, d) = \{n+1, \ldots, d\}$.

Since each of the m groups of scaling terms might occupy different proportions of $\mathbf{\Theta}^{-2}(d)$, we also define the cardinality of the sets $\mathcal{J}(i, d)$:

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

where $c(\mathcal{J}(i, d))$ is assumed to be some polynomial function of the dimension satisfying $\lim_{d\to\infty} c(\mathcal{J}(i, d)) = \infty$.

It will be convenient to rearrange the terms of $\mathbf{\Theta}^{-2}(d)$ so that all the different scaling terms appear at one of the first $n+m$ positions:

$$\Theta^{-2}(d) = \left(\frac{K_1}{d^{\lambda_1}},\dots,\frac{K_n}{d^{\lambda_n}},\frac{K_{n+1}}{d^{\gamma_1}},\dots,\frac{K_{n+m}}{d^{\gamma_m}},\right. \\ \left.\frac{K_{n+1}}{d^{\gamma_1}},\dots,\frac{K_{n+m}}{d^{\gamma_m}},\dots,\frac{K_{n+1}}{d^{\gamma_1}},\dots,\frac{K_{n+m}}{d^{\gamma_m}}\right). \tag{3}$$

This helps to identify each component being studied as $d\to\infty$ without referring to a component that would otherwise be at an infinite position.

Without loss of generality, we assume the first n and the next m scaling terms in (3) to be respectively arranged according to an asymptotic increasing order. If $\preceq$ means "is asymptotically smaller than or equal to", then we have $\theta_1^{-2}(d)\preceq\ldots\preceq\theta_n^{-2}(d)$ and similarly $\theta_{n+1}^{-2}(d)\preceq\ldots\preceq\theta_{n+m}^{-2}(d)$, which respectively implies $-\infty<\lambda_n\le\lambda_{n-1}\le\ldots\le\lambda_1<\infty$ and $-\infty<\gamma_m\le\gamma_{m-1}\le\ldots\le\gamma_1<\infty$. Based on this ordering, the asymptotically smallest scaling term obviously has to be either $\theta_1^{-2}(d)$ or $\theta_{n+1}^{-2}(d)$.

Our goal is to study the limiting distribution of each component forming the d-dimensional Markov process. To this end, we set the scaling term of the target component of interest equal to 1 ($\theta_{i^*}(d)=1$). This adjustment, necessary to the obtention of a nontrivial limiting process, is performed without loss of generality by applying a linear transformation to the target distribution. In particular, when the first component of the chain is studied ($i^*=1$), we set $\theta_1^{-2}(d)=1$ and adjust the other scaling terms accordingly. $\Theta^{-2}(d)$ thus varies according to the component of interest i^* considered.

2.3 The Proposal Distribution and its Scaling

A crucial step in the implementation of Metropolis algorithms is the determination of the optimal form for the proposal scaling as a function of d. Intuitively it makes sense that $\sigma^2(d)$ depends on the asymptotically smallest scaling term in $\Theta^{-2}(d)$. 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.

Moreover, as the dimension of the target increases, more individual moves are 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 d.

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

$$\lim_{d\to\infty}\frac{d^{\lambda_1}}{d^\alpha}<\infty \quad\text{and}\quad \lim_{d\to\infty}\frac{d^{\gamma_i}c(\mathcal{J}(i,d))}{d^\alpha}<\infty,\quad i=1,\dots,m. \tag{4}$$

Therefore, at least one of these $m+1$ limits converges to some positive constant, while the other ones converge to 0. Since the scaling term of the component studied is taken to be 1, then the largest possible form for the proposal scaling is $\sigma^2=\sigma^2(d)=\ell^2$ and so it never diverges as d grows.

By its nature, the Metropolis algorithm is a discrete-time process. Since space (the proposal scaling) is a 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 $d \to \infty$.

Let $\mathbf{Z}^{(d)}(t)$ be the time-t value of the process sped up by a factor of d^α; in particular, $\mathbf{Z}^{(d)}(t) = \left(X_1^{(d)}([d^\alpha t]), \ldots, X_d^{(d)}([d^\alpha t])\right)$, 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 unit time interval. We are now ready to study the limiting comportment of every component of the sequence of processes $\left\{\mathbf{Z}^{(d)}(t), t \geq 0\right\}$ as $d \to \infty$.

3 Optimizing the Sampling Procedure

3.1 Optimal Value for ℓ

We shall now present explicit asymptotic results allowing us to optimize ℓ^2, the constant term of $\sigma^2(d)$. 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 into a statement about efficiency as a function of acceptance rate, as was done in [11].

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 ($\theta_{i^*}(d) = 1$) which, as explained in Section 2.2, might require a linear transformation of $\mathbf{\Theta}^{-2}(d)$.

Theorem 1. *Consider a Metropolis algorithm with proposal distribution* $\mathbf{Y}^{(d)} \sim N\left(\mathbf{x}^{(d)}, \frac{\ell^2}{d^\alpha} I_d\right)$, *where* α *satisfies (4), and applied to a target density as in (1) satisfying the specified conditions on* f, *with* $\theta_j^{-2}(d)$, $j = 1, \ldots, d$ *as in (3) and* $\theta_{i^*}(d) = 1$. *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 (1).*

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)*

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

if and only if

$$\lim_{d \to \infty} \frac{d^{\lambda_1}}{\sum_{j=1}^n d^{\lambda_j} + \sum_{i=1}^m c(\mathcal{J}(i,d)) d^{\gamma_i}} = 0. \tag{5}$$

Here, $\upsilon(\ell) = 2\ell^2 \Phi\left(-\ell\sqrt{E_R}/2\right)$, *and*

$$E_R = \lim_{d \to \infty} \sum_{i=1}^m \frac{c(\mathcal{J}(i,d))}{d^\alpha} \frac{d^{\gamma_i}}{K_{n+i}} \mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^2\right], \tag{6}$$

with $c(\mathcal{J}(i,d))$ *as in (2).*

Intuitively, when none of the target components possesses a scaling term significantly smaller than those of the other components, the limiting process is the same as that found in [11]. Note that the numerator of Condition (5) is based on $\theta_1^{-2}(d)$ only, which is not necessarily the asymptotically smallest scaling term. Technically, we should then also verify if this condition is still satisfied when $\theta_1^{-2}(d)$ is replaced by $\theta_{n+1}^{-2}(d)$; this is however ensured by the term $c(\mathcal{J}(1,d))\,\theta_{n+1}^{2}(d)$ at the denominator.

The function $\upsilon(\ell)$ is sometimes interpreted as the speed measure of the diffusion process. This quantity being proportional to the mixing rate of the algorithm, it then 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 Section 2.1, but where the dependence on the dimension and the proposal scaling are now made explicit. The following corollary introduces the optimal value $\hat{\ell}$ and AOAR leading to greatest efficiency of the Metropolis algorithm.

Corollary 2. *In the setting of Theorem 1 we have* $\lim_{d\to\infty} a(d,\ell) = 2\Phi\left(-\ell\sqrt{E_R}/2\right) \equiv a(\ell)$. *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).*

For a high-dimensional target distribution as defined in Section 2.2 and having no component converging significantly faster than the others, the value ℓ should then be chosen such that the acceptance rate is close to 0.234 in order to optimize the efficiency of the Metropolis algorithm.

Theorem 1 may be used to verify whether the AOAR for sampling from any multivariate normal distribution with covariance matrix Σ is 0.234. Since normal random variables are invariant under orthogonal transformations, we can transform Σ into a diagonal matrix where the eigenvalues of Σ constitute the diagonal elements. The eigenvalues can then be used to verify if Condition (5) 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 Σ with $\sigma_i^2 = 2$, $i = 1,\ldots,d$ and $\sigma_{ij} = 1$, $j \neq i$. The d eigenvalues of Σ are $(d,1,\ldots,1)$ and satisfy Condition (5). For a relatively high-dimensional multivariate normal with such a correlation structure, it is thus optimal to tune the acceptance rate to 0.234. Note however that not all d components mix at the same rate. When studying any of the last $d-1$ components the vector $\mathbf{\Theta}^{-2}(d) = (d,1,\ldots,1)$ is appropriate, so $\sigma^2(d) = \ell^2/d$ and these components thus mix in $O(d)$ iterations. When studying the first component, we need to linearly transform the scaling vector so that $\theta_1^{-2}(d) = 1$. We then use $\mathbf{\Theta}^{-2}(d) = (1, 1/d, \ldots, 1/d)$, so $\sigma^2(d) = \ell^2/d^2$ and this component mixes according to $O\left(d^2\right)$.

Now consider 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 $\sigma_{jk} = 1$, $\forall j \neq k$. Using the d eigenvalues, which are $O(d)$, $O(1/d)$ and 1 with multiplicity $d-2$, we thus conclude that Condition (5) is violated and that 0.234 might not be optimal, even though the distribution is normal (see Theorem 5 of Section 3.2 when dealing with more general $\theta_j(d)$'s).

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 (5) 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 AOAR 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].

3.2 Inhomogeneous Proposal Scaling and Extensions

So far, we have assumed $\sigma^2(d) = \ell^2/d^\alpha$ 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, some components would 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.

The inhomogeneous scheme we adopt is the following: we personalize the proposal scaling of the last $d-n$ components only, implying that the proposal scaling of the first n components is the same as it would have been under the homogeneity assumption. We then treat each of the m groups of scaling terms appearing infinitely often as a different portion of the scaling vector and determine the appropriate α for each group.

In particular, consider $\mathbf{\Theta}^{-2}(d)$ in (3) and let the proposal scaling of X_j be $\sigma_j^2(d) = \ell^2/d^{\alpha_j}$, where $\alpha_j = \alpha$ for $j = 1, \ldots, n$ and α_j is the smallest value such that $\lim_{d\to\infty} c(\mathcal{J}(i,d))\, d^{\gamma_i}/d^{\alpha_j} < \infty$ for $j = n+1, \ldots, d$, $j \in \mathcal{J}(i,d)$. In order to study the component X_{i^*}, we still assume that $\theta_{i^*}(d) = 1$, but we now let $\mathbf{Z}^{(d)}(t) = \mathbf{X}^{(d)}([d^{\alpha_{i^*}} t])$. We have the following result.

Theorem 3. *In the setting of Theorem 1 but with the proposal scaling as just described, the conclusions of Theorem 1 and Corollary 2 are preserved, and E_R is now expressed as*

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

Since the proposal scaling is now adjusted to suit every distinct group of components, each 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 now smaller than with homogeneous proposal scaling. When the proposal scaling of all components was based on α in Section 3.1, 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 $\hat{\ell}^2$.

The conclusions of Section 3.1 also extend to more general target distribution settings. First, we can relax the assumption of equality among the constant terms of $\theta_j^{-2}(d)$ for $j \in \mathcal{J}(i,d)$. In particular, let

$$\begin{aligned} \mathbf{\Theta}^{-2}(d) \quad = \quad & \left(\frac{K_1}{d^{\lambda_1}}, \ldots, \frac{K_n}{d^{\lambda_n}}, \frac{K_{n+1}}{d^{\gamma_1}}, \ldots, \frac{K_{n+c(\mathcal{J}(1,d))}}{d^{\gamma_1}}, \ldots \right. \\ & \left. , \frac{K_{n+\sum_{i=1}^{m-1} c(\mathcal{J}(i,d))+1}}{d^{\gamma_m}}, \ldots, \frac{K_d}{d^{\gamma_m}}\right). \end{aligned} \tag{7}$$

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 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 cannot depend on d, so we 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 assume 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 of the same order, they must necessarily belong to the same of the m groups. We obtain the following result.

Theorem 4. *Consider the setting of Theorem 1, except with $\mathbf{\Theta}^{-2}(d)$ as in (7) 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*

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

if and only if Condition (5) is satisfied. 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.

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 have $c(\mathcal{J}(i,d)) \to \infty$ as $d \to \infty$. We even allow $\left\{\theta_j^{-2}(d), j \in \mathcal{J}(i,d)\right\}$ 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 $\lim_{d\to\infty} \theta_j(d)/\theta_i'(d) = K_j^{-1/2}$ for some reference function $\theta_i'(d)$ and some constant K_j coming from the distribution described for Theorem 4.

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

$$\mathcal{J}(i,d) = \left\{j \in \{1, \ldots, d\}; 0 < \lim_{d\to\infty} \frac{\theta_i'^2(d)}{\theta_j^2(d)} < \infty\right\}, \; i \in \{1, \ldots, m\}. \tag{8}$$

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

$$\begin{aligned} &\lim_{d\to\infty} \theta_1^2(d)\, \sigma_\alpha^2(d) < \infty \quad \text{and} \\ &\lim_{d\to\infty} c(\mathcal{J}(i,d))\, \theta_i'^2(d)\, \sigma_\alpha^2(d) < \infty, \quad i = 1, \ldots, m. \end{aligned} \tag{9}$$

Theorem 5. *Under the setting of Theorem 4, but with proposal scaling* $\sigma^2(d) = \ell^2 \sigma_\alpha^2(d)$ *where* $\sigma_\alpha^2(d)$ *satisfies (9) 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 (5) 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 (8).*

This theorem assumes quite a general form for the scaling terms of the target distribution and allows for a lot of flexibility.

4 Theorems Proofs

We now present the proof of Theorem 1; those of the theorems in Section 3.2 being similar, we just outline the main differences. The proofs are based on Theorem 8.2 of Chapter 4 in [7], which 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. Then, Corollary 8.6 of the same chapter provides further conditions for our sequence of processes to be relatively compact, and thus to reach weak convergence of the stochastic processes themselves. Specifically, it is easily verified that C_c^∞, the space of infinitely differentiable functions on compact support, is an algebra that strongly separates points. Since the algorithm starts in stationarity, $\mathbf{X}^{(d)}(t) \sim \pi \ \forall t > 0$. Using a method similar to the proof of Lemma 7, we show that $\mathrm{E}\left[\left(Gh\left(d, \mathbf{X}^{(d)}\right)\right)^2\right]$ is bounded by some constant for all $d \geq 1$, where G is the generator of the sped up Metropolis algorithm appearing in Section 4.2; this assesses relative compactness.

Our task is then to focus on the $\mathcal{L}^1$ convergence of the generators. To this end, we base our approach on the proof for the Metropolis algorithm case in [10]. Note however that the authors instead prove uniform convergence of generators, which could not be used in the present situation.

The generator is written in terms of an arbitrary test function h, which can usually be any smooth function; in our case, we restrict our attention to functions in C_c^∞. Since the limiting process obtained is a diffusion, then C_c^∞ is a core for the generator by Theorem 2.1 of Chapter 8 in [7], meaning that it is 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: $\mathbf{X}^{(b-a)} = (X_{a+1}, \ldots, X_b)$. The minus sign appearing outside the brackets (e.g. $\mathbf{X}^{(b-a)-}$) means that the component of interest X_{i^*} is excluded. We also use the following notation for conditional expectations: $\mathrm{E}[f(X,Y)\,|X] = \mathrm{E}_Y[f(X,Y)]$. When there is no subscript, the expectation is taken with respect to all random variables included in the expression.

4.1 Restrictions on the Proposal Scaling

We first transform Condition (5) into a statement about the proposal scaling and its parameter α. For this condition to be satisfied, we must equivalently have

$$\lim_{d\to\infty}\frac{K_1}{d^{\lambda_1}}\left(\frac{d^{\lambda_1}}{K_1}+\ldots+\frac{d^{\lambda_n}}{K_n}\right)$$
$$+\lim_{d\to\infty}\frac{K_1}{d^{\lambda_1}}\left(c\left(\mathcal{J}\left(1,d\right)\right)\frac{d^{\gamma_1}}{K_{n+1}}+\ldots+c\left(\mathcal{J}\left(m,d\right)\right)\frac{d^{\gamma_m}}{K_{n+m}}\right)=\infty.$$

Since the first term on the LHS is finite, there is at least one $i\in\{1,\cdots,m\}$ such that $\lim_{d\to\infty}\theta_1^{-2}\left(d\right)c\left(\mathcal{J}\left(i,d\right)\right)\frac{d^{\gamma_i}}{K_{n+i}}=\infty$. Consequently, the choice of α in (4) must be based on one of the groups of scaling terms appearing infinitely often. If we had $\alpha=\lambda_1$, this would mean that $\lim_{d\to\infty}\frac{c(\mathcal{J}(i,d))d^{\gamma_i}}{d^{\alpha}}=\infty$ for all i for which the previous limit was diverging, which contradicts the definition of α. When Condition (5) is satisfied, it thus follows that $\lim_{d\to\infty}d^{\lambda_1}/d^{\alpha}=0$ and $\theta_1^{-2}\left(d\right)$ does not govern α; the parameter α is then strictly greater than 0, no matter which component is under study.

4.2 Proof of Theorem 1

For an arbitrary test function $h\in C_c^{\infty}$, we show that

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

where $Gh\left(d,\mathbf{X}^{(d)}\right)=d^{\alpha}\mathrm{E}_{\mathbf{Y}^{(d)}}\left[\left(h\left(Y_{i^*}\right)-h\left(X_{i^*}\right)\right)\left(1\wedge\frac{\pi\left(d,\mathbf{Y}^{(d)}\right)}{\pi\left(d,\mathbf{X}^{(d)}\right)}\right)\right]$ is the discrete-time generator of the sped up Metropolis algorithm, and $G_Lh\left(X_{i^*}\right)=\upsilon\left(\ell\right)\left[\frac{1}{2}h''\left(X_{i^*}\right)+\frac{1}{2}h'\left(X_{i^*}\right)\left(\log f\left(X_{i^*}\right)\right)'\right]$ is the generator of a Langevin diffusion process with speed measure $\upsilon\left(\ell\right)$ as in Theorem 1.

According to Lemma 7, we have $\lim_{d\to\infty}\mathrm{E}\left[\left|Gh\left(d,\mathbf{X}^{(d)}\right)-\widetilde{G}h\left(d,\mathbf{X}^{(d)}\right)\right|\right]=0$, where

$$\widetilde{G}h\left(d,\mathbf{X}^{(d)}\right)=\frac{1}{2}\ell^2h''\left(X_{i^*}\right)\mathrm{E}_{\mathbf{Y}^{(d)-}}\left[1\wedge e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)}\right]$$
$$+\ell^2h'\left(X_{i^*}\right)\left(\log f\left(X_{i^*}\right)\right)'\mathrm{E}_{\mathbf{Y}^{(d)-}}\left[e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)};\sum_{j=1,j\neq i^*}^{d}\varepsilon\left(d,X_j,Y_j\right)<0\right]$$

and $\varepsilon\left(d,X_j,Y_j\right)$ is as in (10). To prove the theorem, we are thus left to show $\mathcal{L}^1$ convergence of the generator $\widetilde{G}h\left(d,\mathbf{X}^{(d)}\right)$ to that of the Langevin diffusion.

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

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

Since the function h has compact support, then h itself and its derivatives are bounded in absolute value by some constant. As a result, $\mathrm{E}\left[\left|h''\left(X_{i^*}\right)\right|\right]$ 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 expectation on the RHS goes to 0 as $d\to\infty$; we reach the same conclusion for the first expectation of the second term by applying Lemmas 10 and 11.

4.3 Proof of Theorem 4

The main difference with the proof of Theorem 1 happens when working with the m 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 quantity we have to deal with is

$$
\begin{aligned}
&\frac{1}{d^\alpha}\sum_{j\in\mathcal{J}(i,d)}\left(\frac{d}{dX_j}\log\theta_j\left(d\right)f\left(\theta_j\left(d\right)X_j\right)\right)^2 \\
&\quad=\frac{c\left(\mathcal{J}\left(i,d\right)\right)d^{\gamma_i}}{d^\alpha}\left[\frac{1}{c\left(\mathcal{J}\left(i,d\right)\right)}\sum_{j\in\mathcal{J}(i,d)}\left(\frac{d}{dX_j}\log\frac{1}{\sqrt{K_j}}f\left(\frac{X_j}{\sqrt{K_j}}\right)\right)^2\right].
\end{aligned}
$$

By the Weak Law of Large Numbers (WLLN), the term in brackets converges to $b_i\mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^2\right]$. Instead of carrying the term $\theta_{n+i}^2\left(d\right)=d^{\gamma_i}/K_{n+i}$ as before, we thus carry $b_i d^{\gamma_i}$.

4.4 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 $\theta_j\left(d\right)$ for $j\in\mathcal{J}\left(i,d\right)$ are allowed to be different functions of d as long as they are of the same order. Because of this particularity, we have to write $\theta_j\left(d\right)=K_j^{-1/2}\theta_i'\left(d\right)\theta_j^*\left(d\right)/\theta_i'\left(d\right)$, 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 $\theta_{n+i}^2\left(d\right)$ 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.

5 Equivalent Generator and Other Results

5.1 Convergence of an Approximation Term

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

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

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

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

Proof. By Jensen's inequality, $\mathrm{E}\left[|W|\right] \le \sqrt{\mathrm{E}\left[W^2\right]}$. Developing the square and taking the expectation conditional on $\mathbf{X}^{(d)}_{\mathcal{J}(i,d)}$ yield

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

Using changes of variables, we obtain

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

By the WLLN, the term in parentheses on the second line converges in probability to $\mathrm{E}\left[\left(\frac{d^2}{dX^2}\log f\left(X\right)\right)^2\right]$ as $d \to \infty$. Since $d^\alpha > d^{\gamma_i}\sqrt{c\left(\mathcal{J}\left(i,d\right)\right)}$ and the previous expectation is bounded by some constant, the first term converges to 0 as $d \to \infty$. Given that $\theta_{n+i}^2\left(d\right)c\left(\mathcal{J}\left(i,d\right)\right)/d^\alpha$ is $O\left(1\right)$ for at least one $i \in \{1, \ldots, m\}$, we must also show that the term between absolute values converges to 0. From Lemma 12, we know that $f'\left(x\right) \to 0$ as

$x \to \pm\infty$; hence, we have $\mathrm{E}\left[\frac{d^2}{dX_j^2}\log f(X_j) + \left(\frac{d}{dX_j}\log f(X_j)\right)^2\right] = \int f''(x)\,dx = 0$ and as $d \to \infty$, we conclude (by the WLLN) that

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

□

5.2 Convergence to the Equivalent Generator $\widetilde{G}h\left(d,\mathbf{X}^{(d)}\right)$

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

$$\widetilde{G}h\left(d,\mathbf{X}^{(d)}\right) = \frac{1}{2}\ell^2 h''(X_{i^*})\,\mathrm{E}_{\mathbf{Y}^{(d)-}}\left[1 \wedge e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)}\right]$$
$$+\ell^2 h'(X_{i^*})\,(\log f(X_{i^*}))'\,\mathrm{E}_{\mathbf{Y}^{(d)-}}\left[e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)};\ \sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j) < 0\right],$$

where

$$\varepsilon(d,X_j,Y_j) = \log\frac{f(\theta_j(d)\,Y_j)}{f(\theta_j(d)\,X_j)}. \tag{10}$$

If $\alpha > 0$ as defined in (4), then $\lim_{d\to\infty}\mathrm{E}\left[\left|Gh\left(d,\mathbf{X}^{(d)}\right) - \widetilde{G}h\left(d,\mathbf{X}^{(d)}\right)\right|\right] = 0.$

Proof. The proof being similar to that of Lemmas A.2 and A.3 in [10], we shall skip some details. The generator of the sped up Metropolis algorithm can be expressed as

$$Gh\left(d,\mathbf{X}^{(d)}\right) =$$
$$d^\alpha \mathrm{E}_{Y_{i^*}}\left[(h(Y_{i^*}) - h(X_{i^*}))\,\mathrm{E}_{\mathbf{Y}^{(d)-}}\left[1 \wedge e^{\sum_{j=1}^{d}\varepsilon(d,X_j,Y_j)}\right]\right]. \tag{11}$$

We can reexpress the inner expectation using a Taylor expansion of the minimum function with respect to Y_{i^*} and around X_{i^*}. As mentioned in [10] the generator becomes

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

where $g(U_{i^*}) = \varepsilon(X_{i^*}, U_{i^*}) + \sum_{j=1, j\neq i^*}^{d} \varepsilon(d, X_j, Y_j)$, for some $U_{i^*} \in (X_{i^*}, Y_{i^*})$ or (Y_{i^*}, X_{i^*}).

We first note that all expectations computed with respect to $\mathbf{Y}^{(d)-}$ are bounded by 1, $\left|(\log f(U_{i^*}))''\right|$ is bounded by a constant, and $\left|(\log f(U_{i^*}))'\right| \leq \left|(\log f(X_{i^*}))'\right| + K|Y_{i^*} - X_{i^*}|$ for some $K > 0$. Expressing $h(Y_{i^*}) - h(X_{i^*})$ as a three-term Taylor's expansion and using the fact that h has compact support, we can bound the expectations taken with respect to Y_{i^*} and obtain

$$\left|Gh\left(d, \mathbf{X}^{(d)}\right) - \widetilde{G}h\left(d, \mathbf{X}^{(d)}\right)\right| \leq K\left(\frac{\ell^3}{d^{\alpha/2}} + \frac{\ell^4}{d^{\alpha}} + \frac{\ell^5}{d^{3\alpha/2}}\right)\left((\log f(X_{i^*}))'\right)^2$$
$$+ \quad K\left(\frac{\ell^4}{d^{\alpha}} + \frac{\ell^5}{d^{3\alpha/2}} + \frac{\ell^6}{d^{2\alpha}}\right)\left(1 + \left|(\log f(X_{i^*}))'\right|\right) + K\frac{\ell^3}{d^{\alpha/2}} + K\frac{\ell^7}{d^{5\alpha/2}},$$

for some constant $K > 0$. By assumption $\mathrm{E}\left[\left((\log f(X_{i^*}))'\right)^2\right] < \infty$, so it follows that $\mathrm{E}\left[\left|Gh\left(d, \mathbf{X}^{(d)}\right) - \widetilde{G}h\left(d, \mathbf{X}^{(d)}\right)\right|\right]$ converges to 0 as $d \to \infty$. □

6 Volatility and Drift of the Diffusion

6.1 Convergence to an Equivalent Volatility

Lemma 8. *We have*

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

where $\varepsilon(d, X_j, Y_j)$ is as in (10) and

$$z\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right) = \sum_{j=1,j\neq i^*}^{n} \varepsilon(d, X_j, Y_j) + \sum_{i=1}^{m} \sum_{j\in\mathcal{J}(i,d),j\neq i^*} \frac{d}{dX_j} \log f(\theta_j(d) X_j)(Y_j - X_j) - \frac{\ell^2}{2d^{\alpha}} \sum_{i=1}^{m} \sum_{j\in\mathcal{J}(i,d),j\neq i^*} \left(\frac{d}{dX_j} \log f(\theta_j(d) X_j)\right)^2. \quad (12)$$

Proof. Using a Taylor expansion with three terms, we obtain

$$\mathrm{E}_{\mathbf{Y}^{(d)-}}\left[1 \wedge e^{\sum_{j=1,j\neq i^*}^{d} \varepsilon(d,X_j,Y_j)}\right] = \mathrm{E}_{\mathbf{Y}^{(d)-}}\left[1 \wedge \exp\left\{\sum_{j=1,j\neq i^*}^{n} \varepsilon(d, X_j, Y_j) + \sum_{i=1}^{m} \sum_{j\in\mathcal{J}(i,d),j\neq i^*} \left[\frac{d}{dX_j} \log f(\theta_j(d) X_j)(Y_j - X_j) + \frac{1}{2}\frac{d^2}{dX_j^2} \log f(\theta_j(d) X_j)(Y_j - X_j)^2 + \frac{1}{6}\frac{d^3}{dU_j^3} \log f(\theta_j(d) U_j)(Y_j - X_j)^3\right]\right\}\right], \quad (13)$$

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

By the triangle's inequality, the Lipschitz property of the function $1 \wedge e^x$ (see Proposition 2.2 in [11]) and noticing that the first two terms of the function $z\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right)$ cancel out with the first two terms of the exponential function in (13), we get

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

By Lemma 6, the RHS converges in probability to 0 as $d \to \infty$. We then apply the Bounded Convergence Theorem to complete the proof of the lemma. □

6.2 Simplified Expression for the Equivalent Volatility

Lemma 9. *If Condition (5) is satisfied, then*

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

where $z\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right)$ and E_R are as in (12) and (6) respectively.

Proof. For each group of components whose scaling term appears infinitely often in the limit, i.e. for $i = 1, \ldots, m$ let

$$R_i\left(d, \mathbf{x}^{(d)-}_{\mathcal{J}(i,d)}\right) = \frac{1}{d^\alpha} \sum_{j\in\mathcal{J}(i,d), j\neq i^*} \left(\frac{d}{dx_j} \log f\left(\theta_j(d)\, x_j\right)\right)^2. \tag{14}$$

Since $(Y_j - X_j)\,|X_j \sim$ *iid* $N\left(0, \ell^2/d^\alpha\right)$ for $j = 1, \cdots, d$, then

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

Applying Proposition 2.4 in [11] allows us to obtain an expression in terms of $\Phi(\cdot)$, the *cdf* of a standard normal random variable

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

We note that $E_R > 0$ since there is at least one $i \in \{1, \ldots, m\}$ such that $\lim_{d \to \infty} c(\mathcal{J}(i,d)) d^{\gamma_i}/d^{\alpha} > 0$. Using Propositions 13 and 14, and then applying Slutsky's Theorem and the Continuous Mapping Theorem, we conclude that $\exp\left(\sum_{j=1, j \neq i^*}^{n} \varepsilon(d, X_j, Y_j)\right) \to_p 1$ and

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

Since $\mathrm{E}_{\mathbf{Y}^{(d-n)-}}\left[1 \wedge e^{z\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right)}\right]$ is positive and bounded by 1, we use the Bounded Convergence Theorem to assert that $\mathrm{E}\left[1 \wedge e^{z\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right)}\right] \to_p 2\Phi\left(-\ell\sqrt{E_R}/2\right)$; we complete the proof of the lemma by reapplying the Bounded Convergence Theorem. □

6.3 Convergence to an Equivalent Drift

Lemma 10. *We have*

$$\lim_{d\to\infty} \mathrm{E}_{\mathbf{X}^{(d)-}}\left[\left|\mathrm{E}_{\mathbf{Y}^{(d)-}}\left[e^{\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)}; \sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j) < 0\right] - \mathrm{E}_{\mathbf{Y}^{(d)-}}\left[e^{z\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)}; z\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right) < 0\right]\right|\right] = 0, \tag{15}$$

where $\varepsilon(d, X_j, Y_j)$ and $z\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right)$ are as in (10) and (12) respectively.

Proof. First, let $T(x) = e^x \mathbf{1}_{(x<0)}$,

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

and

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

We shall show that $A\left(d, \mathbf{Y}^{(d)-}, \mathbf{X}^{(d)-}\right) \left| \mathbf{X}^{(d)-} \right. \to_p 0$, and then use this result to prove convergence of expectations.

Similar to the proof of Lemma A.7 in [10], we have

$$\begin{aligned} &\mathrm{P}_{\mathbf{Y}^{(d)-}}\left(\left|A\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right| \geq \delta(d)\right) \leq \\ &\quad \mathrm{P}_{\mathbf{Y}^{(d)-}}\left(\left|\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j) - z\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right| \geq \delta(d)\right) \\ &\quad + \mathrm{P}_{\mathbf{Y}^{(d)-}}\left(-\delta(d) < z\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right) < \delta(d)\right). \end{aligned} \tag{16}$$

By Markov's inequality and the proof of Lemma 8, the first term on the RHS is bounded by

$$\frac{1}{\delta(d)}\mathrm{E}_{\mathbf{Y}^{(d)-}}\left[\left|\sum_{j=1,j\neq i^*}^{d}\varepsilon(d,X_j,Y_j)-z\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\right|\right]\leq\sqrt{\delta(d)}\to_p 0$$

as $d\to\infty$. Using conditioning and the proof of Lemma 9, the second term on the RHS becomes

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

Using the convergence results developed in the proof of Lemma 9 along with the fact that $\delta(d)\to_p 0$ as $d\to\infty$ and the Bounded Convergence Theorem, we deduce that the previous expression converges in probability to 0. Therefore, $A\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)\left|\mathbf{X}^{(d)-}\right.\to_p 0$ and (15) follows by reapplying the Bounded Convergence Theorem twice. □

6.4 Simplified Expression for the Equivalent Drift

Lemma 11. *If Condition (5) is satisfied, then*

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

where the functions $\varepsilon(d,X_j,Y_j)$ *and* $z\left(d,\mathbf{Y}^{(d)-},\mathbf{X}^{(d)-}\right)$ *are as in (10) and (12) respectively.*

Proof. The proof is similar to that of Lemma 9, the only difference lying in the fact that (Proposition 2.4 in [11])

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

□

7 Discussion

The theorems in this paper basically extend the *iid* work of [11] 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 are similar to those in [11], since the AOARs are identical; the sole difference lies in the optimal scaling values themselves. Condition (5), which says that no target component converges significantly faster than the others, ensures that the process asymptotically behaves as in the *iid* case. This work thus partially answers Open Problem #3 of [14].

These results can also be used to determine, for virtually any correlated multivariate normal target distribution, whether or not 0.234 is optimal. Contrarily to what seemed to be a common belief, multivariate normal distributions do not always adopt a conventional limiting behavior and there exist cases where the AOAR is significantly smaller than 0.234 (see [1]).

It was shown in the *iid* case that although asymptotic, the results are 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 $\lambda_j < \alpha$, otherwise Condition (5) would be violated) then the optimal acceptance rate converges extremely slowly to 0.234 from above. For instance, suppose that $\mathbf{\Theta}^{-2}(d) = \left(d^{-\lambda}, 1, \ldots, 1\right)$ with $\lambda < 1$. The proposal scaling is then $\sigma^2(d) = \ell^2/d$ and the closer to 1 is λ, the slower is the convergence of the optimal acceptance rate to 0.234. In fact, for a multivariate normal target with $\lambda = 0.75$, simulations show that d must be as big as $200,000$ for the optimal acceptance rate to be reasonably close to 0.234; they also show that for $\alpha - \lambda \geq 0.5$, the asymptotic results are accurate in relatively small dimensions, just as in the *iid* case. Detailed examples and simulation studies illustrating the results introduced in this paper and in [1] are presented in [2].

Appendix

Lemma 12. *Let f be a C^2 probability density function (pdf). If $(\log f(x))'$ is Lipschitz continuous, then $f'(x) \to 0$ as $x \to \pm\infty$.*

Proof. The asymptotic behavior of a C^2 *pdf* as $x \to \pm\infty$ can be one of three things: (1) $f(x) \to 0$, $f'(x) \to 0$; (2) $f(x) \to 0$, $f'(x) \nrightarrow 0$; (3) $f(x) \nrightarrow 0$, $f'(x) \nrightarrow 0$. We prove that in cases (2) and (3), $(\log f(x))'$ is not Lipschitz continuous, which implies that (1) is the only possible option.

(2) $f(x) \to 0$, $f'(x) \nrightarrow 0$: Since $f \to 0$, then $\forall \epsilon > 0$, $\exists x_0(\epsilon) \in \mathbf{R}$ such that $\forall x \geq x_0(\epsilon)$, $f(x) < \epsilon$. Since $f' \nrightarrow 0$, then $\forall \epsilon > 0$, $\exists x^* \geq x_0(\epsilon) + 1$ such that $|f'(x^*)| > \limsup |f'|/2$. Because f is C^2, then $\forall 0 < \epsilon < \limsup |f'|/2$, $\exists y$ with $|x^* - y| \leq 1$ such that $|f'(y)| = \epsilon$. Now, choose y^* to be the value y which minimizes $|x^* - y|$, but such that $f(y^*) > f(x^*)$. Given $0 < \epsilon < \limsup |f'|/2$, we then have

$$\sup_{x,y \in \mathbf{R}, x \neq y} \frac{\left|\frac{f'(x)}{f(x)} - \frac{f'(y)}{f(y)}\right|}{|x-y|} \geq \frac{\left|\frac{|f'(x^*)|}{f(x^*)} - \frac{|f'(y^*)|}{f(y^*)}\right|}{1} \geq \left|\frac{\limsup |f'|/2 - \epsilon}{\epsilon}\right|.$$

Since this is true for all $0 < \epsilon < \limsup |f'| /2$, then the Lipschitz continuity assumption is violated.

(3) $f(x) \nrightarrow 0$, $f'(x) \nrightarrow 0$: Since f is continuous, positive, and $\int f = 1$, then $\forall \epsilon > 0$, $\exists x_0(\epsilon) \in \mathbf{R}$ such that $f(x) < \epsilon$ for $x \geq x_0(\epsilon)$, except on a set A_ϵ of Lebesgue measure $\lambda(A_\epsilon) < \epsilon$. Since $(-\infty, \epsilon)$ is an open set, then $B = \{x \in \mathbf{R} : f(x) < \epsilon\}$ must be open as well; $A_\epsilon = B^c \cap [x_0(\epsilon), \infty)$ is then formed of closed intervals over which $f(x) \geq \epsilon$.

Since $f \nrightarrow 0$, then $\forall \epsilon > 0$, $\exists$ an interval $[x(\epsilon), y(\epsilon)]$ in A_ϵ where the maximum value reached by f over this interval ($h(\epsilon)$ say) is such that $h(\epsilon) > \limsup |f| /2$. There might be many values in the interval for which $f(x) = h(\epsilon)$, but all these values satisfy $f'(x) = 0$. Since $f(x(\epsilon)) = f(y(\epsilon)) = \epsilon$, then $\sup_{x\in\mathbf{R}} f'(x) \geq \frac{h(\epsilon)-\epsilon}{y(\epsilon)-x(\epsilon)} > \frac{h(\epsilon)-\epsilon}{\epsilon}$. Hence, $\sup_{x\in\mathbf{R}} \frac{f'(x)}{f(x)} > \frac{h(\epsilon)-\epsilon}{\epsilon h(\epsilon)}$ and since this is true $\forall \epsilon > 0$, then $\sup_{x\in\mathbf{R}} \frac{f'(x)}{f(x)} = \infty$. Given $\epsilon > 0$, we take y to be one of the points in $[x(\epsilon), y(\epsilon)]$ such that $f(y) = h(\epsilon)$ and $f'(y) = 0$. We then have

$$\sup_{x,y\in\mathbf{R}, x\neq y} \frac{\left|\frac{f'(x)}{f(x)} - \frac{f'(y)}{f(y)}\right|}{|x-y|} \geq \sup_{x\in\mathbf{R}} \frac{\left|\frac{f'(x)}{f(x)} - 0\right|}{|x(\epsilon) - y(\epsilon)|} > \sup_{x\in\mathbf{R}} \frac{\left|\frac{f'(x)}{f(x)} - 0\right|}{\epsilon} = \infty,$$

and we realize that the Lipschitz continuity assumption is violated. Note that in cases (2) and (3), we have considered the case where $x \to \infty$; we can repeat a similar reasoning for the case where $x \to -\infty$. □

Proposition 13. *Let $\varepsilon(d, X_j, Y_j)$, $j = 1, \ldots, n$ be as in (10). If $\lambda_j < \alpha$, then $\varepsilon(d, X_j, Y_j) \to_p 0$.*

Proof. By Taylor's Theorem, we have for some $U_j \in (X_j, Y_j)$ or (Y_j, X_j)

$$\begin{aligned}\mathrm{E}[|\varepsilon(d, X_j, Y_j)|] = \mathrm{E}\Big[\Big|&(\log f(\theta_j(d) X_j))'(Y_j - X_j) + \\ &\frac{1}{2}(\log f(\theta_j(d) X_j))''(Y_j - X_j)^2 + \frac{1}{6}(\log f(\theta_j(d) U_j))'''(Y_j - X_j)^3\Big|\Big].\end{aligned}$$

Applying changes of variables and using the fact that $\left|(\log f(X))''\right|$ and $\left|(\log f(U))'''\right|$ are bounded by a constant, we get for some $K > 0$

$$\mathrm{E}[|\varepsilon(d, X_j, Y_j)|] \leq \ell \frac{d^{\lambda_j/2}}{d^{\alpha/2}} K \mathrm{E}\left[\left|(\log f(X))'\right|\right] + \left(\ell^2 \frac{d^{\lambda_j}}{d^\alpha} + \ell^3 \frac{d^{3\lambda_j/2}}{d^{3\alpha/2}}\right) K.$$

By assumption, $\mathrm{E}\left[\left|(\log f(X))'\right|\right]$ 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 Markov's inequality and find that for all $\epsilon > 0$, $\mathrm{P}(|\varepsilon(d, X_j, Y_j)| \geq \epsilon) \leq \mathrm{E}[|\varepsilon(d, X_j, Y_j)|]/\epsilon \to 0$ as $d \to \infty$. □

Proposition 14. *Let $R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)$ be as in (14), with $i \in \{1, \ldots, m\}$. We 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 (6).*

Proof. The expectation of each variable satisfies $\mathrm{E}\left[R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\right] = \frac{c(\mathcal{J}(i,d))}{d^\alpha}\frac{d^{\gamma_i}}{K_{n+i}}\mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^2\right]$. By independence between the X_j's and using the fact that $\mathrm{Var}\,(X) \leq \mathrm{E}\left[X^2\right]$, we obtain

$$\mathrm{Var}\left(\sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\right) \leq \sum_{i=1}^{m} \frac{1}{d^{2\alpha}}\frac{d^{2\gamma_i}}{K^2_{n+i}} c\left(\mathcal{J}\left(i,d\right)\right) \mathrm{E}\left[\left(\frac{f'\left(X\right)}{f\left(X\right)}\right)^4\right].$$

By assumption, $\mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^4\right]$ is finite and since $c\left(\mathcal{J}\left(i,d\right)\right) d^{2\gamma_i} < d^{2\alpha}$, the variance converges to 0 as $d \to \infty$. To conclude the proof, we use Chebychev's inequality and find that $\forall \epsilon > 0$, $\mathrm{P}\left(\left|\sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right) - E_R\right| \geq \epsilon\right) \leq \frac{1}{\epsilon^2}\mathrm{Var}\left(\sum_{i=1}^{m} R_i\left(d, \mathbf{X}^{(d)-}_{\mathcal{J}(i,d)}\right)\right) \to 0$ as $d \to \infty$. □

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, proof-reading by Stephen B. Connor, as well as constructive comments by the anonymous referee; those have led to an improved paper.

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 and stochastic systems. *Statist. Sci.* **10**, 3-66.

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

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

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

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

[9] 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.

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

[11] 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.

[12] 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.

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

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

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