Optimal Scaling for Random walk Metropolis on spherically constrained target densities

Neal, Peter and Roberts, Gareth

2007

MIMS EPrint: **2006.417**

ISSN 1749-9097

Optimal Scaling for Random walk Metropolis on spherically constrained target densities

Peter Neal* and Gareth Roberts†‡

June 29, 2007

Abstract

We consider the problem of optimal scaling of the proposal variance for multidimensional Random walk Metropolis (RWM) algorithms. It is well known, for a wide range of continuous target densities, that the optimal scaling of the proposal variance leads to an average acceptance rate of 0.234. Therefore a natural question is, do similar results hold for target densities which have discontinuities? In the current work, we answer in the affirmative for a class of spherically constrained target densities. Even though the acceptance probability is more complicated than for continuous target densities, the optimal scaling of the proposal variance again leads to an average acceptance rate of 0.234.

AMS 2000 subject classification. Primary 60F05; secondary 65C05.

Keywords: Random walk Metropolis algorithm, Markov chain Monte Carlo, optimal scaling, spherical distributions.

1 Introduction

The Random walk Metropolis (RWM) algorithm is one of the most widely used Markov chain Monte Carlo (MCMC) algorithms. The RWM algorithms popularity is due to the fact that it is easy to implement and its generic nature. Therefore it is often seen as the default MCMC algorithm when more model specific algorithms do not readily present themselves. However the RWM algorithms generic nature can

*School of Mathematics, University of Manchester, Sackville Street, Manchester, M60 1QD , P.Neal-2@manchester.ac.uk

†Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, g.o.roberts@lancaster.ac.uk

‡New address: Dept of Statistics, University of Warwick, Coventry, CV4 7AL, UK, Gareth.O.Roberts@warwick.ac.uk

be its downfall and it is important that the proposal variance is carefully chosen to construct an efficient algorithm. If the proposal variance is too small, then the RWM algorithm converges slowly since all of the increments are small. Alternatively, if the proposal variance is too large, the RWM will reject too high a proportion of proposed moves.

The question of optimal scaling of RWM algorithms for d-dimensional target distributions has received considerable attention. A number of heuristic, 'rules of thumb' have been proposed, see Besag and Green (1993) and Besag *et al.* (1995). However, in Roberts *et al.* (1997) theoretical guidelines were obtained by considering a sequence of d-dimensional target distributions as $d \to \infty$. These guidelines although asymptotic have been shown to be practically useful for relatively low dimensions such as $d = 10$. Moreover the guidelines provided by Roberts *et al.* (1997) are easy to implement and summed up in the following statement from Roberts *et al.* (1997), page 113.

Tune the proposal variance so that the average acceptance rate is roughly 1/4. (1.1)

In Roberts *et al.* (1997), iid product densities were considered. Subsequent papers have shown that (1.1) holds in a range of situations, see Breyer and Roberts (2000), Roberts and Rosenthal (2001), Neal and Roberts (2006) and Bédard (2007). All these papers consider continuous target densities. Therefore the following question is posed; does (1.1) hold for discontinuous target densities? A partial answer is given in this paper, in that, we show that (1.1) holds when the target distribution is subjected to a global (spherical) constraint on the components. In a subsequent paper, Neal *et al.* (2007), we show that (1.1) does not hold for target distributions with local discontinuities, that is, where the discontinuities are given in terms of individual components as opposed to a global condition.

The paper is structured as follows. In Section 2, the target distribution to be considered is introduced. In Section 3, RWM on the d-dimensional uniform hypersphere is considered. In particular, we focus on the limiting behaviour of movements both in the radial component of the hypersphere and individual components. The analysis is similar to Roberts *et al.* (1997), thus allowing for direct comparisons with the results there in. However, variation in the radial component, and hence the acceptance probability, leads to more involved arguments than those required in Roberts *et al.* (1997). In Section 4, extending the results of Section 3 to more general target distributions is discussed. This begins with a detailed comparison with Roberts *et al.* (1997) and is followed by analysis of constrained Gaussian random variables for which explicit results can be derived. In Section 5, a simulation study is presented to demonstrate the general applicability of the results given in the previous sections. Also some limitations of the rule (1.1) are discussed with examples of densities where the addition of the spherical constraint to the density leads to an average acceptance probability of 0.234 being sub-optimal. Finally, in Section 6 a brief summary of the results is given.

2 Target densities

For $d \geq 1$, we consider the optimal scaling of the proposal variance for target distributions of the form:

$$\pi_d(\mathbf{x}^d) \propto \begin{cases} \prod_{i=1}^d f(x_i) & \text{if } \frac{1}{d}\sum_{i=1}^d x_i^2 \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

The spherical constraint is chosen so that in the limit, as $d \to \infty$, each of the components of $\mathbf{X}^d \sim \pi_d(\cdot)$ have non-trivial marginal distributions. Note that without the spherical constraint (2.1) is the product density considered in Roberts *et al.* (1997).

In Section 3, we consider the d-dimensional uniform hypersphere, that is, the special case of (2.1) where $f(x) = 1$ ($x \in \mathbb{R}$). The behaviour of the RWM algorithm in this case is indicative of the RWM behaviour for more general target densities. In Section 4, we let $f(\cdot) \sim N(0, \lambda)$ for $\lambda > 0$, with the d-dimensional uniform hypersphere as the special limiting case when $\lambda \to \infty$. Complications encountered when considering more general choices of $f(\cdot)$ are also discussed in Section 4.

The RWM algorithm is described below. For $t \geq 0$ and $i \geq 1$, let $Z_{t,i}$ be independent and identically distributed according to $Z \sim N(0,1)$. For $d \geq 1$, $1 \leq i \leq d$, $t \geq 0$ and $l > 0$, let $\sigma_d = l/\sqrt{d}$ and

$$Y_{t+1,i}^d = X_{t,i}^d + \sigma_d Z_{t,i}.$$

Then if $\frac{1}{d}\sum_{i=1}^d (Y_{t+1,i}^d)^2 \leq 1$, we accept the proposed move with probability $1 \wedge \prod_{i=1}^d f(Y_{t+1,i}^d)/f(X_{t,i}^d)$. If the move is accepted, we set $\mathbf{X}_{t+1}^d = \mathbf{Y}_{t+1}^d$. Otherwise, we reject the move and set $\mathbf{X}_{t+1}^d = \mathbf{X}_t^d$.

The stationary distribution of $\mathbf{X}_t^d$ is given by (2.1). Each of the components of $\mathbf{X}_t^d$ are identically distributed and exchangeable. Therefore we shall focus on the first two components $X_{\cdot,1}^d$ and $X_{\cdot,2}^d$. In particular, we show that the movements in the first two components are asymptotically independent. For $t \geq 0$, let $R_t^d = \left(\frac{1}{d}\sum_{i=1}^d (X_{t,i}^d)^2\right)^{\frac{1}{2}}$ denote the (normalised) radius. A key point to note is that for the uniform hypersphere and the constrained Gaussian distribution the acceptance probability is totally determined by the radius of $\mathbf{Y}_{t+1}^d$. Therefore in both cases we begin by studying the behaviour of the radial component before analysing $X_{\cdot,1}^d$ and $X_{\cdot,2}^d$ in detail.

3 Hypersphere

In this section we consider the uniform hypersphere case, that is,

$$\pi_d(\mathbf{x}^d) \propto \begin{cases} 1 & \text{if } \frac{1}{d}\sum_{i=1}^d x_i^2 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

Movements in the radial component are analysed in Section 3.1 with the analysis of $(X^d_{\cdot,1}, X^d_{\cdot,2})$ presented in Section 3.2.

3.1 Radial Component

For large d, the majority of the mass of the hypersphere is located close to the surface (radius equal to 1). Under stationarity, R_0^d has cumulative distribution function $F_d(r) = r^d$ $(0 \le r \le 1)$. Therefore rather than consider the asymptotic behaviour of R_t^d as $d \to \infty$, it will be convenient to consider $B_t^d = -d \log R_t^d$, where for all $d \ge 1$, $B_0^d \sim Exp(1)$.

Fix $l > 0$. We shall assume that l is fixed for the remainder of this section. For $t \ge 0$, let the Markov chain B have the following transition kernel,

$$B_{t+1} = \begin{cases} B_t - \tilde{Z}_t & \text{if } B_t - \tilde{Z}_t > 0 \\ B_t & \text{otherwise,} \end{cases} \tag{3.2}$$

where $\tilde{Z}_t \sim N\left(\frac{l^2}{2}, l^2\right)$. The Markov chain $\{B_t\}$ is a random walk on the positive half line with stationary distribution $Exp(1)$. Geometric ergodicity of $\{B_t\}$ is easily verified using a Foster-Lyapunov drift condition. The proof is similar to that given in Meyn and Tweedie (1993), Section 16.1.3, pages 394–5, and hence, the details are omitted.

Before showing that B is the limiting process of B^d we introduce some preliminary results. Let

$$\tilde{B}^d(\mathbf{x}^d) = -\frac{d}{2} \log\left(\frac{1}{d}\sum_{i=1}^d (x_i^d)^2\right).$$

Then for $\alpha, \gamma > 0$, let

$$F_d^{(\alpha,\gamma)} = \{\mathbf{x}^d; \max_{1 \le i \le d} |x_i^d| \le d^\alpha\} \cap \{\mathbf{x}^d; 0 \le \tilde{B}^d(\mathbf{x}^d) \le \gamma \log d\}.$$

We then have the following trivial result which will enable us, for $\alpha > 0$ and $\gamma > 1$, to restrict attention to $\mathbf{X}_t^d \in F_d^{(\alpha,\gamma)}$.

Lemma 3.1 *For all $\alpha > 0$ and $\gamma > 1$,*

$$d\mathbb{P}(\mathbf{X}_0^d \notin F_d^{(\alpha,\gamma)}) \to 0 \qquad d \to \infty. \tag{3.3}$$

Proof. Fix $\alpha > 0$ and $\gamma > 1$. Note that

$$d\mathbb{P}(\mathbf{X}_0^d \notin F_d^{(\alpha,\gamma)}) \leq d\mathbb{P}(\max_{1\leq i\leq d} |X_{0,i}^d| > d^\alpha) + d\mathbb{P}(\tilde{B}^d(\mathbf{X}^d) > \gamma \log d). \tag{3.4}$$

The components of $\mathbf{X}_0^d$ are exchangeable, and so, the first term on the righthandside of (3.4) is bounded as follows

$$d\mathbb{P}(\max_{1\leq i\leq d} |X_{0,i}^d| > d^\alpha) \leq d^2\mathbb{P}(|X_{0,1}^d| > d^\alpha). \tag{3.5}$$

Therefore since $X_{0,1}^d$ has probability density function,

$$g_d(x) = \begin{cases} \frac{\Gamma(d/2+1)}{\pi^{d/2}(\sqrt{d})^d}\left(\frac{\pi^{d-1/2}(\sqrt{d-x^2})^{d-1}}{\Gamma((d-1)/2+1)}\right) & -\sqrt{d} \leq x \leq \sqrt{d} \\ 0 & \text{otherwise,} \end{cases}$$

it is straightforward to show that the righthandside of (3.5) converges to 0 as $d \to \infty$.

The latter term on the righthandside of (3.4) converges to 0 as $d \to \infty$, since for all $t \geq 0$, $\tilde{B}^d(\mathbf{X}_t^d) \sim Exp(1)$. □

Lemma 3.2 *For all $0 < \alpha < \frac{1}{4}$, $\gamma > 1$, $k \in \mathbb{N}$ and $t \geq 0$, if $\mathbf{X}_t^d \in F_d^{(\alpha,\gamma)}$, then*

$$\frac{l}{\sqrt{d}}\sum_{i=k}^d X_{t,i}^d Z_{t,i} + \frac{l^2}{2d}\sum_{i=k}^d Z_{t,i}^2 \stackrel{D}{=} \tilde{Z}_t + \epsilon_t^d \qquad \text{as } d \to \infty, \tag{3.6}$$

where $\tilde{Z}_t \sim N(l^2/2, l^2)$ and for any $\delta < 1 - 4\alpha$, $d^\delta \epsilon_t^d \xrightarrow{p} 0$ as $d \to \infty$.

Proof. Let $\mathbf{X}_t^d = \mathbf{w}^d \in F_d^{(\alpha,\gamma)}$. We prove the result for $t = 1$, the general result follows similarly.

Let $\theta \in \mathbb{R}$. Then since $\max_{1\leq j\leq d} |w_j^d| \leq d^\alpha$,

$$\begin{aligned} \mathbb{E}\left[\exp\left(i\theta\left\{\sigma_d \sum_{j=1}^d w_j^d Z_{1,j} + \frac{\sigma_d^2}{2}\sum_{j=1}^d Z_{1,j}^2\right\}\right)\right] &= \prod_{j=1}^d \mathbb{E}\left[\exp\left(i\theta\left\{\sigma_d w_j^d Z_{1,j} + \frac{\sigma_d^2}{2}Z_{1,j}^2\right\}\right)\right] \\ &= \prod_{j=1}^d \left(1 + i\theta\frac{l^2}{2d} - \frac{\theta^2 l^2}{2d}(w_j^d)^2 + O(d^{4\alpha-2})\right) \\ &= \exp\left(\sum_{j=1}^d \left\{i\theta\frac{l^2}{2d} - \frac{\theta^2 l^2}{2d}(w_j^d)^2 + O(d^{4\alpha-2})\right\}\right) \\ &= \exp\left(i\theta\frac{l^2}{2} - \theta^2\frac{l^2}{2}\frac{1}{d}\sum_{j=1}^d (w_j^d)^2 + O(d^{4\alpha-1})\right). \end{aligned} \tag{3.7}$$

Note that since $\mathbf{w}^d \in F_d^{(\alpha,\gamma)}$, we have that

$$1 - 2\frac{\gamma \log d}{d} \leq \frac{1}{d}\sum_{j=3}^{d}(w_j^d)^2 \leq 1.$$

Therefore it follows from (3.7) that

$$\sigma_d \sum_{j=3}^{d} w_j^d Z_{1,j} + \frac{\sigma_d^2}{2}\sum_{j=3}^{d} Z_{1,j}^2 \stackrel{D}{=} \tilde{Z}_1 + \epsilon_1^d$$

where for any $\delta < 1 - 4\alpha$, $d^\delta \epsilon_1^d \stackrel{p}{\longrightarrow} 0$ as $d \to \infty$. □

Lemmas 3.1 and 3.2 are stronger than are required for analysing the radial component but are needed for the analysis of the individual components in Section 3.2. We now turn our attention to the main results for the radial component.

Theorem 3.3 *For all $b \in \mathbb{R}^+$ and $T \in \mathbb{N}$,*

$$B_T^d | B_0^d = b \stackrel{D}{\longrightarrow} B_T | B_0 = b \qquad \text{as } d \to \infty. \tag{3.8}$$

Proof. We prove the result for $T = 1$. The result for general $T \in \mathbb{N}$ follows straightforwardly since $B_\cdot^d$ and $B_\cdot$ are Markovian.

For $d \geq 1$ and $t \geq 1$, let

$$S_{t+1}^d = -\frac{d}{2}\log\left(\frac{1}{d}\sum_{i=1}^{d}(Y_{t+1,i}^d)^2\right),$$

then

$$B_{t+1}^d = \begin{cases} S_{t+1}^d & \text{if } S_{t+1}^d > 0 \\ B_t^d & \text{otherwise.} \end{cases}$$

Firstly, note that

$$\begin{aligned}
\sum_{i=1}^{d}(Y_{1,i}^d)^2 &= \left(\mathbf{X}_0^d + \frac{l}{\sqrt{d}}\mathbf{Z}_0^d\right)^T \left(\mathbf{X}_0^d + \frac{l}{\sqrt{d}}\mathbf{Z}_0^d\right) \\
&= d(R_0^d)^2 + 2\frac{l}{\sqrt{d}}(\mathbf{Z}_0^d)^T \mathbf{X}_0^d + \frac{l^2}{d}(\mathbf{Z}_0^d)^T \mathbf{Z}_0^d \\
&= d(R_0^d)^2 + 2\frac{l}{\sqrt{d}}\sum_{i=1}^{d} X_{0,i}^d Z_{0,i} + \frac{l^2}{d}\sum_{i=1}^{d} Z_{0,i}^2.
\end{aligned}$$

By Lemma 3.2,

$$\begin{aligned}
d\left\{\exp\left(-\frac{2}{d}S_{t+1}^d\right) - \exp\left(-\frac{2}{d}B_t^d\right)\right\} &= \sum_{i=1}^{d}(Y_{t+1,i}^d)^2 - d(R_t^d)^2 \\
&\stackrel{D}{\longrightarrow} 2\tilde{Z}_t \qquad \text{as } d \to \infty.
\end{aligned} \tag{3.9}$$

For all $t \geq 0$,

$$\left| d \left\{ \exp\left(-\frac{2}{d} S_{t+1}^d\right) - \exp\left(-\frac{2}{d} B_t^d\right) \right\} + 2(S_{t+1}^d - B_t^d) \right| \stackrel{p}{\longrightarrow} 0 \qquad \text{as } d \to \infty,$$

and so, by Billingsley (1968), Theorem 4.1,

$$-(S_{t+1}^d - B_t^d) \stackrel{D}{\longrightarrow} \tilde{Z}_t \qquad \text{as } d \to \infty.$$

Therefore for all $b \geq 0$,

$$S_{t+1}^d | B_t^d = b \stackrel{D}{\longrightarrow} S_{t+1} | B_t = b \qquad \text{as } d \to \infty, \tag{3.10}$$

where

$$S_{t+1} = B_t - \tilde{Z}_t. \tag{3.11}$$

Since $\tilde{Z}_t$ is continuous, (3.8) follows from (3.10) and (3.11). □

Theorem 3.3 shows that the radial component mixes in $O(1)$ iterations. However, for studying the movement in individual components we shall require the following result.

Lemma 3.4 *For any $\beta > 0$, $\gamma > 1$ and for all $b_d \in [0, \gamma \log d]$,*

$$B_{[d^\beta]}^d | B_0^d = b_d \stackrel{D}{\longrightarrow} \tilde{B} \sim Exp(1) \qquad \textit{as } d \to \infty. \tag{3.12}$$

Proof. Fix $\beta, \epsilon > 0$, $\gamma > 1$, $\zeta \in \mathbb{R}^+$ and let $C = [0, \zeta]$.

Let $W_d = \min_{1 \leq i \leq d} \{B_i^d \in C\}$. Then since $B_\cdot^d$ has negative drift, it is trivial to show that

$$\mathbb{P}(W_d > [d^\beta/2]) \to 0 \qquad \text{as } d \to \infty. \tag{3.13}$$

Since $\{B_t\}$ is geometrically ergodic, there exists $T \in \mathbb{N}$ such that

$$|\{B_T | B_0 \in C\} - \tilde{B}|_{TV} < \frac{\epsilon}{2}, \tag{3.14}$$

see Meyn and Tweedie (1993) page 354, Theorem 15.0.1. However, for all $b \in C$ and $x \in \mathbb{R}$,

$$\begin{aligned} & |\mathbb{P}(B_T^d \leq x | B_0^d = b) - \mathbb{P}(\tilde{B} \leq x)| \\ \leq \quad & |\mathbb{P}(B_T^d \leq x | B_0^d = b) - \mathbb{P}(B_T \leq x | B_0 = b)| + |\mathbb{P}(B_T \leq x | B_0 = b) - \mathbb{P}(\tilde{B} \leq x)|. \end{aligned} \tag{3.15}$$

By (3.8) and (3.14), respectively, the two terms on the righthandside of (3.15) are bounded by $\epsilon/2$ for all sufficiently large d. Therefore since $[d^\beta/2] \to \infty$ as $d \to \infty$, it follows that for all sufficiently large d,

$[d^\beta/2] \geq T$, and so, by the Markov property

$$B^d_{[d^\beta/2]} | B^d_0 = b \xrightarrow{D} \tilde{B} \quad \text{as } d \to \infty. \tag{3.16}$$

The lemma follows from (3.13) and (3.16). □

3.2 Individual Components

We are now in position to consider the movement in any of the components. Since the components are exchangeable but not independent we shall focus upon components 1 and 2.

For $t \geq 0$ and $d \geq 1$, let $\mathbf{U}^d_t = \left(X^d_{[dt],1}, X^d_{[dt],2} \right)$.

Theorem 3.5 *For all $d \geq 1$, let $\mathbf{X}^d_0$ be distributed according to $\pi_d(\cdot)$, where for $\mathbf{x}^d \in \mathbb{R}^d$,*

$$\pi_d(\mathbf{x}^d) \propto \begin{cases} 1 & \text{if } \frac{1}{d}\sum_{i=1}^d (x_i^d)^2 \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.17}$$

Then, as $d \to \infty$,

$$\mathbf{U}^d \Rightarrow \mathbf{U} \quad \text{as } d \to \infty,$$

where $\mathbf{U}_\cdot = (U_{\cdot,1}, U_{\cdot,2})$, $U_{0,i} \sim N(0,1)$ $(i = 1,2)$ and $\mathbf{U}$ satisfies the Langevin SDE

$$d\mathbf{U}_t = s(l)^{1/2} d\mathbf{B}_t - \frac{s(l)}{2} \mathbf{U}_t \, dt \tag{3.18}$$

with $s(l) = 2l^2 \Phi(-l/2)$. (Note that Φ and ϕ denote the cdf and pdf of a standard normal random variable, respectively.)

Thus the limiting process $\mathbf{U}$ is a bivariate Ornstein-Uhlenbeck process with independent components and each component having stationary distribution $N(0,1)$. Hence in the limit as $d \to \infty$ any pair of components are (asymptotically) independent. Furthermore, the statement of Theorem 3.5 is essentially identical to the statement of Roberts *et al.* (1997), Theorem 1.1. In particular, the speed measure of the diffusion is of the same form. Thus letting $a_d(l)$ denote the $\pi_d(\cdot)$ average acceptance rate of the d-dimensional RWM, we have the following Corollary which mirrors Roberts *et al.* (1997), Corollary 1.2.

Corollary 3.6

$$\lim_{d \to \infty} a_d(l) = a(l) = 2\Phi\left(-\frac{l}{2} \right).$$

$s(l)$ is maximised by $\hat{l} = 2.38$ with $a(\hat{l}) = 0.234$.

We proceed by introducing the notation and results needed to prove Theorem 3.5. Fix $0 < \alpha, \beta, \tau < \frac{1}{16}$ and $\gamma > 1$ with $\alpha + \beta < \tau$. For $t \geq 0$, let $\mathbf{W}_t^{d,\tau} = \mathbf{X}_{tk_d^\tau}^d$, where $k_d^\tau = [d^\tau]$. Thus the $\mathbf{W}_\cdot^{d,\tau}$ processes are the $\mathbf{X}_\cdot^d$ processes observed at time-points $0, k_d^\tau, 2k_d^\tau, \ldots$.

Let G_d^τ be the (discrete-time) generator of $\mathbf{W}^{d,\tau}$, and let $V \in C_c^\infty$ (the space of infinitely differentiable functions on compact support) be an arbitrary test function of the first two components only. Thus

$$\begin{aligned}
G_d^\tau V(\mathbf{w}^d) &= d^{1-\tau}\mathbb{E}\left[V(\mathbf{W}_1^{d,\tau}) - V(\mathbf{W}_0^{d,\tau})|\mathbf{W}_0^{d,\tau} = \mathbf{w}^d\right] \\
&= d^{1-\alpha}\mathbb{E}\left[V(\mathbf{X}_{k_d^\tau}^d) - V(\mathbf{X}_0^d)|\mathbf{X}_0^d = \mathbf{w}^d\right] \\
&= d^{1-\tau}\sum_{i=0}^{k_d^\tau-1}\mathbb{E}\left[V(\mathbf{X}_{i+1}^d) - V(\mathbf{X}_i^d)|\mathbf{X}_0^d = \mathbf{w}^d\right] \\
&= d^{1-\tau}\sum_{i=0}^{k_d^\tau-1}\mathbb{E}\left[(V(\mathbf{Y}_{i+1}^d) - V(\mathbf{X}_i^d))\left\{1 \wedge \frac{\pi_d(\mathbf{Y}_{i+1}^d)}{\pi_d(\mathbf{X}_i^d)}\right\}\middle|\mathbf{X}_0^d = \mathbf{w}^d\right] \qquad (3.19)
\end{aligned}$$

The generator G of the two-dimensional Ornstein-Uhlenbeck process described in Theorem 3.5, for an arbitrary test function $V \in C_c^\infty$, is given by

$$GV(w_1, w_2) = s(l)\sum_{i=1}^{2}\left\{\frac{1}{2}\frac{\partial^2}{\partial w_i^2}V(w_1, w_2) - \frac{w_i}{2}\frac{\partial}{\partial w_i}V(w_1, w_2)\right\}. \qquad (3.20)$$

By Ethier and Kurtz (1986), Chapter 4, Corollary 8.7 and Lemma 3.1, we can restrict attention to $\mathbf{X}_t^d \in F_d^{(\alpha,\gamma)}$. (*i.e.* $\mathbf{X}_t^d$ stays close to the boundary of the hypersphere, and none of the components are excessively large.) The aim will therefore be to show that,

$$\sup_{\mathbf{w}^d \in F_d^{(\alpha,\gamma)}} |G_d V(\mathbf{w}^d) - GV(w_1, w_2)| \to 0 \quad \text{as } d \to \infty. \qquad (3.21)$$

Before proving (3.21) rigorously we give an outline of the arguments used in the proof. The acceptance probability is a function of the radius which mixes in $O(1)$ iterations. On the other hand, any single component mixes in $O(d)$ iterations. Thus the acceptance probability is mixing much faster than any of the individual components. Therefore for any $0 < \beta < \alpha$, the radial component has 'forgotten' its starting value after $[d^\beta]$ iterations (see Lemma 3.4), whereas any given component barely moves in $[d^\beta]$ iterations. Furthermore, over $[d^\beta]$ iterations approximately $a_d(l)[d^\beta]$ proposed moves will be accepted.

For $b \geq 0$ and $u_1, u_2 \in \mathbb{R}$, let

$$h(b, u_1, u_2) = \frac{l^2}{2}\sum_{i=1}^{2}\left\{\Phi\left(\frac{1}{l}\left\{b - \frac{l^2}{2}\right\}\right)\frac{\partial^2}{\partial u_i^2}V(u_1, u_2) - \frac{1}{l}u_i\phi\left(\frac{1}{l}\left\{b - \frac{l^2}{2}\right\}\right)\frac{\partial}{\partial u_i}V(u_1, u_2)\right\}. \qquad (3.22)$$

Lemma 3.7 *For all* $\mathbf{w}^d \in F_d^{(\alpha,\gamma)}$,

$$\mathbb{E}\left[(V(\mathbf{Y}_1^d) - V(\mathbf{X}_0^d)) \left\{ 1 \wedge \frac{\pi_d(\mathbf{Y}_1^d)}{\pi_d(\mathbf{X}_0^d)} \right\} \middle| \mathbf{X}_0^d = \mathbf{w}^d \right] = \frac{1}{d} h(b_d, w_1, w_2) + o(d^{-5/4}), \tag{3.23}$$

where $b_d = -\frac{d}{2} \log \left(\frac{1}{d} \sum_{i=1}^d w_i^2 \right)$.

Therefore there exists $K < \infty$ *such that for all* $d \geq 1$ *and* $\mathbf{w}^d \in F_d^{(\alpha,\gamma)}$,

$$d\mathbb{E}\left[(V(\mathbf{Y}_1^d) - V(\mathbf{X}_0^d)) \left\{ 1 \wedge \frac{\pi_d(\mathbf{Y}_1^d)}{\pi_d(\mathbf{X}_0^d)} \right\} \middle| \mathbf{X}_0^d = \mathbf{w}^d \right] \leq K d^\alpha. \tag{3.24}$$

Proof. Let

$$A_d(\mathbf{Y}_1^d) = \begin{cases} 1 & \text{if } \left\{ \frac{1}{d} \sum_{i=1}^d (Y_{1,i}^d)^2 \right\}^{1/2} \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, letting $\mathbf{Y}_-^d = (Y_3^d, Y_4^d, \ldots, Y_d^d)$,

$$\begin{aligned}
& \mathbb{E}\left[(V(\mathbf{Y}_1^d) - V(\mathbf{X}_0^d)) \left\{ 1 \wedge \frac{\pi_d(\mathbf{Y}_1^d)}{\pi_d(\mathbf{X}_0^d)} \right\} \middle| \mathbf{X}_0^d = \mathbf{w}^d \right] \\
= \; & \mathbb{E}\left[(V(\mathbf{Y}_1^d) - V(\mathbf{X}_0^d)) A_d(\mathbf{Y}_1^d) | \mathbf{X}_0^d = \mathbf{w}^d \right] \\
= \; & \mathbb{E}_{Y_1^d, Y_2^d} \left[(V(\mathbf{Y}_1^d) - V(\mathbf{X}_0^d)) \mathbb{E}_{\mathbf{Y}_-^d} [A_d(\mathbf{Y}_1^d) | \mathbf{X}_0^d = \mathbf{w}^d, Y_1^d, Y_2^d] | \mathbf{X}_0^d = \mathbf{w}^d \right].
\end{aligned}$$

Concentrating first on the inner expectation. Note that for $1 \leq i \leq d$, $Y_{1,i}^d = w_i + \sigma_d Z_{1,i}$, where $Z_{1,i} \sim N(0,1)$. Thus,

$$\begin{aligned}
& \mathbb{E}_{\mathbf{Y}_-^d} [A_d(\mathbf{Y}_1^d) | \mathbf{X}_0^d = \mathbf{w}^d, Z_{1,1} = z_1, Z_{1,2} = z_2] \\
= \; & \mathbb{P}\left(\frac{1}{d} \sum_{i=1}^d (Y_{1,i}^d)^2 \leq 1 \,\middle|\, \mathbf{X}_0^d = \mathbf{w}^d, Y_1^d = w_1 + \sigma_d z_1, Y_2^d = w_2 + \sigma_d z_2 \right) \\
= \; & \mathbb{P}\left(d(R_0^d)^2 + 2\sigma_d \left(w_1 z_1 + w_2 z_2 + \sum_{i=3}^d w_i Z_{1,i} \right) + \sigma_d^2 \left(z_1^2 + z_2^2 + \sum_{i=3}^d Z_{1,i}^2 \right) \leq d \middle| \mathbf{X}_0^d = \mathbf{w}^d \right) \\
= \; & \mathbb{P}\left(d - 2b_d + o(d^{-3/4}) + 2\sigma_d \left(w_1 z_1 + w_2 z_2 + \sum_{i=3}^d w_i Z_{1,i} \right) + \sigma_d^2 \left(z_1^2 + z_2^2 + \sum_{i=3}^d Z_i^2 \right) \leq d \middle| \mathbf{X}_0^d = \mathbf{w}^d \right).
\end{aligned} \tag{3.25}$$

Therefore by Lemma 3.2, (3.7)

$$\begin{aligned}
& \mathbb{P}\left(d - 2b_d + o(d^{-3/4}) + 2\sigma_d \left(w_1 z_1 + w_2 z_2 + \sum_{i=3}^d w_i Z_{1,i} \right) + \sigma_d^2 \left(z_1^2 + z_2^2 + \sum_{i=3}^d Z_{1,i}^2 \right) \leq d \middle| \mathbf{X}_0^d = \mathbf{w}^d \right) \\
= \; & \mathbb{P}\left(\tilde{Z}_1 + \epsilon_d \leq b_d - o(d^{-3/4}) - \sigma_d(w_1 z_1 + w_2 z_2) + \frac{\sigma_d^2}{2}(z_1^2 + z_2^2) \right) \\
= \; & \int_{-\infty}^{\infty} f_{\epsilon_d}(x) \Phi\left(\frac{1}{l} \left\{ b_d - \frac{l^2}{2} - \sigma_d(w_1 z_1 + w_2 z_2) - x + o(d^{-3/4}) \right\} \right) dx \\
= \; & \Phi\left(\frac{1}{l} \left\{ b_d - \frac{l^2}{2} \right\} \right) - \frac{1}{l} \sigma_d (w_1 z_1 + w_2 z_2) \phi\left(\frac{1}{l} \left\{ b_d - \frac{l^2}{2} \right\} \right) + o(d^{-3/4}),
\end{aligned} \tag{3.26}$$

using a Taylor series expansion.

Also by Taylor's Theorem,

$$\begin{aligned}
& V(\mathbf{Y}_1^d) - V(\mathbf{X}_0^d) \\
= \; & \frac{l}{\sqrt{d}}\left(z_1\frac{\partial}{\partial w_1}V(\mathbf{w}^d) + z_2\frac{\partial}{\partial w_2}V(\mathbf{w}^d)\right) \\
& +\frac{l^2}{2d}\left(z_1^2\frac{\partial^2}{\partial w_1^2}V(\mathbf{w}^d) + z_2^2\frac{\partial^2}{\partial w_2^2}V(\mathbf{w}^d) + z_1 z_2\frac{\partial^2}{\partial w_1 w_2}V(\mathbf{w}^d)\right) + o(d^{-5/4}). \qquad (3.27)
\end{aligned}$$

Therefore it follows from (3.26) and (3.27) that

$$\begin{aligned}
& \mathbb{E}\left[(V(\mathbf{Y}_1^d) - V(\mathbf{X}_0^d))A_d(\mathbf{Y}_1^d)|\mathbf{X}_0^d = \mathbf{w}^d\right] \\
= \; & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(z_1)\phi(z_2)\left\{\frac{l}{\sqrt{d}}\left(z_1\frac{\partial}{\partial w_1}V(\mathbf{w}^d) + z_2\frac{\partial}{\partial w_2}V(\mathbf{w}^d)\right)\right. \\
& \left. +\frac{l^2}{2d}\left(z_1^2\frac{\partial^2}{\partial w_1^2}V(\mathbf{w}^d) + z_2^2\frac{\partial^2}{\partial w_1^2}V(\mathbf{w}^d) + z_1 z_2\frac{\partial^2}{\partial w_1 w_2}V(\mathbf{w}^d)\right) + o(d^{-5/4})\right\} \\
\times \; & \left\{\Phi\left(\frac{1}{l}\left\{b_d - \frac{l^2}{2}\right\}\right) - \frac{1}{\sqrt{d}}(w_1 z_1 + w_2 z_2)\phi\left(\frac{1}{l}\left\{b_d - \frac{l^2}{2}\right\}\right) + o(d^{-3/4})\right\} dz_2 dz_1 \\
= \; & \frac{l^2}{2d}\sum_{i=1}^{2}\left\{\Phi\left(\frac{1}{l}\left\{b_d - \frac{l^2}{2}\right\}\right)\frac{\partial^2}{\partial w_i^2}V(w_1,w_2) - \frac{1}{l}w_i\phi\left(\frac{1}{l}\left\{b_d - \frac{l^2}{2}\right\}\right)\frac{\partial}{\partial w_i}V(w_1,w_2)\right\} + o(d^{-5/4}) \\
= \; & \frac{1}{d}h(b_d, w_1, w_2) + o(d^{-5/4}), \qquad (3.28)
\end{aligned}$$

and (3.23) is proved.

Finally (3.24) follows straightforwardly from (3.28) since $V \in C_c^\infty$ and for $\mathbf{w}^d \in F_d^{(\alpha,\gamma)}$, $|w_1|, |w_2| \leq d^\alpha$.
□

Lemma 3.8 *For any* $\mathbf{w}^d \in F_d^{(\alpha,\gamma)}$ *and for any sequence of positive integers* $\{c_d\}$ *such that* $[d^\beta] \leq c_d \leq [d^\tau]$,

$$d\mathbb{E}\left[(V(\mathbf{Y}_{c_d+1}^d) - V(\mathbf{X}_{c_d}^d))\left\{1 \wedge \frac{\pi_d(\mathbf{Y}_{c_d+1}^d)}{\pi_d(\mathbf{X}_{c_d}^d)}\right\}\middle|\mathbf{X}_0^d = \mathbf{w}^d\right] \to \int_0^\infty h(b, w_1, w_2)e^{-b}\, db.$$

Proof. Fix the sequence $\{c_d\}$ such that for all $d \geq 1$, $[d^\beta] \leq c_d \leq [d^\alpha]$. By Lemma 3.7, for $\mathbf{w}^d \in F_d^{(\alpha,\gamma)}$,

$$\begin{aligned}
& d\mathbb{E}\left[(V(\mathbf{Y}_{c_d+1}^d) - V(\mathbf{X}_{c_d}^d))A_d(\mathbf{Y}_{c_d+1}^d)|\mathbf{X}_0^d = \mathbf{w}^d\right] \\
= \; & d\int_0^\infty\left\{\int \mathbb{E}\left[(V(\mathbf{Y}_{c_d+1}^d) - V(\mathbf{X}_{c_d}^d))A_d(\mathbf{Y}_{c_d+1}^d)|\mathbf{X}_{c_d}^d = \mathbf{u}^d\right] g_{c_d}(\mathbf{u}^d|\mathbf{w}^d, b_d(\mathbf{u}^d) = b)\, d\mathbf{u}^d\right\} f_{c_d}^d(b|\mathbf{w}^d)\, db \\
= \; & \int_0^\infty\left\{\int \{h_d(b_d(\mathbf{u}^d), u_1, u_2) + o(d^{-1/4})\} g_{c_d}^d(\mathbf{u}^d|\mathbf{w}^d, b_d(\mathbf{u}^d) = b)\, d\mathbf{u}^d\right\} f_{c_d}^d(b|\mathbf{w}^d)\, db, \qquad (3.29)
\end{aligned}$$

where $f_{c_d}^d(\cdot|\mathbf{w}^d)$ and $g_{c_d}^d(\cdot|\mathbf{w}^d)$ denote the pdfs of $B_{c_d}^d$ and $\mathbf{X}_{c_d}^d$, respectively, given that $\mathbf{X}_0^d = \mathbf{w}^d$.

For any $\epsilon > 0$ and $|u_1 - w_1|, |u_2 - w_2| < O(d^{-\epsilon})$, it follows by Taylor's Theorem that

$$h(b_d(\mathbf{u}^d), u_1, u_2) \quad = \quad h(b_d(\mathbf{u}^d), w_1, w_2) + O(d^{-\epsilon}). \tag{3.30}$$

For any $i \geq 1$, by the triangle inequality,

$$|X^d_{c_d,i} - X^d_{0,i}| \leq \sigma_d \sum_{j=1}^{c_d} |Z_{j,i}|.$$

Let $\epsilon = \frac{1}{2} - 2\tau$. By Markov's inequality,

$$d\mathbb{P}\left(\sigma_d \sum_{j=1}^{c_d} |Z_{j,i}| > d^{-\epsilon}\right) \to 0 \quad \text{ as } d \to \infty. \tag{3.31}$$

Therefore by (3.30) and (3.31), it follows from (3.29) that

$$\begin{aligned} & d\mathbb{E}\left[(V(\mathbf{Y}^d_{c_d+1}) - V(\mathbf{X}^d_{c_d}))A_d(\mathbf{Y}^d_{c_d+1})|\mathbf{X}^d_0 = \mathbf{w}^d\right] \\ = & \int_0^\infty \left\{ \int \{h_d(b, w_1, w_2) + O(d^{-\epsilon})\} g_{c_d}(\mathbf{u}^d|\mathbf{w}^d, b_d(\mathbf{u}^d) = b)\, d\mathbf{u}^d \right\} f^d_{c_d}(b|\mathbf{w}^d)\, db \\ = & \int_0^\infty h(b, w_1, w_2) f^d_{c_d}(b|\mathbf{w}^d)\, db + \delta_d, \end{aligned} \tag{3.32}$$

where $\delta_d \to 0$ as $d \to \infty$.

By Lemma 3.4, for all $\mathbf{w}^d \in F_d^{(\alpha,\gamma)}$, $B^d_{c_d}|\mathbf{X}^d_0 = \mathbf{w}^d \stackrel{D}{\longrightarrow} \tilde{B}$ as $d \to \infty$. Furthermore, for all $w_1, w_2 \in \mathbb{R}$, $h(\cdot, w_1, w_2)$ is bounded. Therefore the righthand-side of (3.32) converges to

$$\int_0^\infty h(b, w_1, w_2) f(b)\, db = \int_0^\infty h(b, w_1, w_2) e^{-b}\, db \quad \text{ as } d \to \infty$$

and the lemma is proved. □

Lemma 3.9

$$\int_0^\infty h(b, w_1, w_2) e^{-b}\, db = s(l) \sum_{i=1}^2 \left\{ \frac{1}{2} \frac{\partial^2}{\partial w_i^2} V(w_i) - \frac{w_i}{2} \frac{\partial}{\partial w_i} V(w_i) \right\}.$$

Proof. The lemma follows straightforwardly since

$$\int_0^\infty e^{-b} \Phi\left(\frac{b}{l} - \frac{l}{2}\right)\, db = 2\Phi\left(-\frac{l}{2}\right)$$

and

$$\frac{1}{l} \int_0^\infty e^{-b} \phi\left(\frac{b}{l} - \frac{l}{2}\right)\, db = \Phi\left(-\frac{l}{2}\right),$$

as required. □

Proof of Theorem 3.5. By Ethier and Kurtz (1986), Chapter 4, Corollary 8.7 to prove the theorem it is sufficient to show that,

$$\sup_{\mathbf{w}^d \in F_d^{(\alpha,\gamma)}} |G_d^\tau V(\mathbf{w}^d) - GV(w_1, w_2)| \to 0 \quad \text{as } d \to \infty.$$

It follows trivially from Corollary 3.8 and Lemma 3.9, that for all $\mathbf{w}^d \in F_d^{(\alpha,\gamma)}$,

$$\begin{aligned} G_d^\tau V(\mathbf{w}^d) &= d^{1-\tau}\mathbb{E}\left[V(\mathbf{W}_1^{d,\tau}) - V(\mathbf{W}_0^{d,\tau})|\mathbf{W}_0^{d,\tau} = \mathbf{w}^d\right] \\ &= d^{1-\tau}\sum_{i=0}^{[d^\beta]}\mathbb{E}\left[V(\mathbf{X}_{i+1}^d) - V(\mathbf{X}_i^d)|\mathbf{X}_0^d = \mathbf{w}^d\right] \\ &\quad + d^{1-\tau}\sum_{i=[d^\beta]+1}^{[d^\tau]}\mathbb{E}\left[V(\mathbf{X}_{i+1}^d) - V(\mathbf{X}_i^d)|\mathbf{X}_0^d = \mathbf{w}^d\right]. \end{aligned} \tag{3.33}$$

Since $\alpha + \beta < \tau$, the first term on the righthand-side of (3.33) converges to 0 as $d \to \infty$ by (3.24). The second term converges to $GV(w_1, w_2)$ as $d \to \infty$ by Lemmas 3.8 and 3.9.

Since C_c^∞ separates points (see Ethier and Kurtz (1986), page 113), the theorem follows from (3.33) by Ethier and Kurtz (1986), Chapter 4, Corollary 8.7 provided the compact containment condition holds for $\{\mathbf{U}_\cdot^d\}$. This is easily verified using the proof of Neal and Roberts (2006), Theorem 3.1 □

4 Constrained Gaussian distributions

4.1 Introduction

In Section 3, we have considered the special case where $\pi_d(\cdot)$ is uniformly distributed over the d-dimensional hypersphere with radius $\sqrt{d}$. We shall in this section consider extensions of the results of Section 3. However, we begin by comparing the results obtained so far with previous analysis in Roberts *et al.* (1997) giving reasons for restricting attention to constrained Gaussian random variables.

The acceptance probability for the d-dimensional hypersphere is totally determined by the radial component which mixes in $O(1)$ iterations. In Roberts *et al.* (1997) it was shown that for independent and identically distributed product densities $\pi_d(\mathbf{x}^d) = \prod_{i=1}^d f(x_i)$ where $f(\cdot) = \exp(g(\cdot))$ is the pdf of X, the acceptance probability of a move is determined by $\frac{1}{d-1}\sum_{i=2}^d g'(x_i)^2$ and $\frac{1}{d-1}\sum_{i=2}^d g''(x_i)$. Furthermore, under stationarity, subject to mild moment conditions upon $g'(\cdot)$ and $g''(\cdot)$, $\frac{1}{d-1}\sum_{i=2}^d g'(X_{t,i})^2$ and $\frac{1}{d-1}\sum_{i=2}^d g''(X_{t,i})$ are essentially constant for all $t \geq 0$. That is, with sufficiently high probability for large d, the acceptance probability is contained within $\left(\Phi(-l\sqrt{I}/2) - \epsilon_d, \Phi(-l\sqrt{I}/2) + \epsilon_d\right)$ where

$I \equiv \mathbb{E}_f[g'(X)^2]$ and $\epsilon_d \to 0$ as $d \to \infty$. Thus attention in Roberts *et al.* (1997) can be restricted to the movement of individual components. Therefore the movement in the radial component of the hypersphere is a complication not encountered in Roberts *et al.* (1997).

For more general target densities than the hypersphere the acceptance probability is more complicated than a 0-1 indicator. In particular, for $\pi_d(\cdot)$ given by (2.1), (the constrained version of Roberts *et al.* (1997) (1.1)),

$$\begin{aligned} 1 \wedge \frac{\pi_d(\mathbf{Y}_1^d)}{\pi_d(\mathbf{x}^d)} &= 1_{\{d^{-1}\sum_i (Y_{1,i}^d)^2 \leq 1\}} \left\{ 1 \wedge \prod_{i=1}^d \frac{f(Y_{1,i}^d)}{f(x_i^d)} \right\} \\ &= 1_{\{b_d - \sigma_d^2/2 \sum_i Z_{1,i}^2 - \sigma_d \sum_i x_i^d Z_{1,i} + o(d^{-1/4}) > 0\}} \\ &\quad \times \left\{ 1 \wedge \exp\left(\sigma_d \sum_{i=1}^d g'(x_i) Z_{1,i} + \frac{1}{2}\sigma_d^2 \sum_{i=1}^d g''(x_i) Z_{1,i}^2 + o(d^{-1/4}) \right) \right\} \end{aligned} \tag{4.1}$$

with $\mathbf{X}_0^d = \mathbf{x}^d$ and $b_d = -\frac{d}{2}\log\left(\sum_i (x_i^d)^2\right)$. Therefore (4.1) is a hybrid of the acceptance probability of Section 3 and Roberts *et al.* (1997), and the joint distribution of

$$\left(\sigma_d \sum_{i=1}^d x_i^d Z_{t,i} + \frac{\sigma_d^2}{2} \sum_{i=1}^d Z_{t,i}^2, \sigma_d \sum_{i=1}^d g'(x_i^d) Z_{t,i} + \frac{\sigma_d^2}{2} \sum_{i=1}^d g''(x_i^d) Z_{t,i}^2 \right)$$

needs to be studied. In Roberts *et al.* (1997), it is shown that $\frac{1}{d-1}\sum_{i=2}^d g'(X_{t,i})^2 \approx I$, but such arguments do not readily extend to the current target density due to the dependencies in the components of $\mathbf{X}_t^d$ induced by the constraint.

Progress can be made when $f(x) = \frac{1}{\sqrt{2\pi\lambda}}\exp(-x^2/2\lambda)$ $(x \in \mathbb{R})$, *i.e.* $f(\cdot) \sim N(0, \lambda)$. In this case

$$\left(\sigma_d \sum_{i=1}^d x_i^d Z_{t,i} + \frac{\sigma_d^2}{2} \sum_{i=1}^d Z_{t,i}^2, \sigma_d \sum_{i=1}^d g'(x_i^d) Z_{t,i} + \frac{\sigma_d^2}{2} \sum_{i=1}^d g''(x_i^d) Z_{t,i}^2 \right) = Q_t^d(1, 1/\lambda)$$

where

$$Q_t^d (= Q_t^d(\mathbf{x}^d)) = \sigma_d \sum_{i=1}^d x_i^d Z_{t,i} + \frac{\sigma_d^2}{2} \sum_{i=1}^d Z_{t,i}^2.$$

Thus the acceptance probability is determined by Q_t^d and $B_t^d = -\frac{d}{2}\log\left(\sum_i (X_{t,i}^d)^2\right)$.

Without any constraint, if $X_1, X_2, \ldots$ are independent and identically distributed according to $X \sim N(0, \lambda)$, then

$$\frac{1}{d}\sum_i X_i^2 \xrightarrow{a.s.} \lambda \qquad \text{as } d \to \infty. \tag{4.2}$$

Therefore with the constraint there are three cases to consider $\lambda < 1$, $\lambda = 1$ and $\lambda > 1$. For $\lambda < 1$, the constraint $\frac{1}{d}\sum_i X_i^2 < 1$ is redundant, and so, Roberts *et al.* (1997), Theorem 1.1 holds. Furthermore, the constraint is redundant for any $Y \sim f(\cdot)$ for which $\mathbb{E}[Y^2] < 1$. Thus we restrict attention to the cases where the constraint is important. In particular, we shall focus on $\lambda > 1$ where the results mirror those of the hypersphere. Note that the hypersphere is the limiting case as $\lambda \to \infty$. Finally, the case $\lambda = 1$ is more intricate with a different scaling of the radial component. In particular, the mixing of the (scaled) radial component is $O(d)$ and the methodology required for dealing with this is very different to that used here. As a consequence, we shall consider the case $\lambda = 1$ elsewhere.

4.2 Radial component

The analysis is very similar to section 3.1, and so, only an outline of the argument is given.

For $t \geq 0$, let the Markov chain B have transition kernel,

$$B_{t+1} = \begin{cases} B_t - \tilde{Z}_t & \text{with probability } 1 \wedge \exp\left(-\tilde{Z}_t/\lambda\right) \text{ if } B_t - \tilde{Z}_t > 0 \\ B_t & \text{otherwise.} \end{cases}$$

where $\tilde{Z}_t \sim N(l^2/2, l^2)$. It is straightforward by studying the balance equation to show that B_t has stationary distribution $Exp(\mu_\lambda)$ where $\mu_\lambda = \frac{\lambda-1}{\lambda}$. The geometric ergodicity of $\{B_t\}$ can be easily verified using Foster-Lyapunov drift criteria, as for the uniform hypersphere in Section 3.1.

Theorem 4.1 *For all $b > 0$,*

$$B_1^d | B_0^d = b \xrightarrow{D} B_1 | B_0 = b \qquad as\ d \to \infty. \tag{4.3}$$

For any $\beta > 0$, $\gamma > 1/\mu_\lambda$ and for all $b_d \in [0, \gamma \log d]$,

$$B_{[d^\beta]}^d | B_0^d = b \xrightarrow{D} \tilde{B}_\lambda \sim Exp(\mu_\lambda) \qquad as\ d \to \infty. \tag{4.4}$$

Proof. The proofs of (4.3) and (4.4) are essentially identical to the proofs of Theorem 3.3 and Lemma 3.4, respectively, and therefore the details are omitted. □

We conclude our brief analysis of the radial component by noting that in the conditions of Theorem 4.1, $\gamma > 1/\mu_\lambda$ replaces $\gamma > 1$ for the hypersphere. This is necessary for (3.3) to hold for the constrained Gaussian. We can then utilise the sets $\{F_d^{(\alpha,\gamma)}\}$ as before when considering the movements of the individual components.

4.3 Individual Components

For $t \geq 0$ and $d \geq 1$, let $\mathbf{U}_t^d = (X_{[dt],1}^d, X_{[dt],2}^d)$. Theorem 4.2 is virtually identical to Theorem 3.5.

Theorem 4.2 *Suppose that there exists $\lambda > 1$ such that $f(\cdot) \sim N(0, \lambda)$. For all $d \geq 1$, let $\mathbf{X}_0^d$ be distributed according to $\pi_d(\cdot)$ (2.1), where for $\mathbf{x}^d \in \mathbb{R}^d$,*

$$\mathbf{U}^d \Rightarrow \mathbf{U} \qquad \text{as } d \to \infty,$$

where $\mathbf{U}_{\cdot} = (U_{1,\cdot}, U_{2,\cdot})$, $U_{0,i} \sim N(0,1)$ $(i = 1, 2)$ and $U_{1,\cdot}$ and $U_{2,\cdot}$ are independent Ornstein-Uhlenbeck processes with $U_{i,0} \sim N(0,1)$ $(i = 1, 2)$ and $\mathbf{U}$ satisfies the Langevin SDE

$$d\mathbf{U}_t = s(l)^{1/2} d\mathbf{B}_t - \frac{s(l)}{2} \mathbf{U}_t \, dt \tag{4.5}$$

with $s(l) = 2l^2 \Phi(-l/2)$.

The proof of Theorem 4.2 is similar to the proof of Theorem 3.5. Whilst some of the calculations are a little more involved, the essentials of the proof are the same. Therefore we give an outline of the proof only highlighting the salient points.

Proposition 4.3 *For any $c \in \mathbb{R}$ and for $Z \sim N(\mu, \sigma^2)$,*

$$\mathbb{E}\left[1_{\{Z<c\}}\{1 \wedge \exp(-Z)\}\right] = \Phi\left(-\frac{\mu}{\sigma}\right) + \exp\left(\frac{\sigma^2}{2} - \mu\right)\left\{\Phi\left(\frac{c-\mu}{\sigma} + \sigma\right) - \Phi\left(\sigma - \frac{\mu}{\sigma}\right)\right\} \tag{4.6}$$

and for $c > 0$,

$$\mathbb{E}\left[1_{\{Z<c\}} \exp(-Z); Z > 0\right] = \exp\left(\frac{\sigma^2}{2} - \mu\right)\left\{\Phi\left(\frac{c-\mu}{\sigma} + \sigma\right) - \Phi\left(\sigma - \frac{\mu}{\sigma}\right)\right\}. \tag{4.7}$$

Lemma 4.4 *For any $\lambda > 1$ and for $\mathbf{X}_0^d = \mathbf{x}^d$,*

$$\begin{aligned}
\mathbb{E}_{\mathbf{Y}_1^{d-}}\left[1 \wedge \frac{\pi_d(\mathbf{Y}_1^d)}{\pi_d(\mathbf{x}^d)}\right] &= \Phi\left(-\frac{l}{2}\right) + \exp\left(-\frac{l^2}{2}\frac{\mu_\lambda}{\lambda}\right)\left\{\Phi\left(\frac{b_d}{l} + \frac{l}{\lambda} - \frac{l}{2}\right) - \Phi\left(\frac{l}{\lambda} - \frac{l}{2}\right)\right\} \\
&\quad + \sum_{i=1}^{2} -\sigma_d x_i z_i \exp\left(-\frac{l^2}{2}\frac{\mu_\lambda}{\lambda}\right)\left\{\frac{1}{l}\phi\left(\frac{b_d}{l} + \frac{l}{\lambda} - \frac{l}{2}\right)\right. \\
&\quad \left. + \frac{1}{\lambda}\left\{\Phi\left(\frac{b_d}{l} + \frac{l}{\lambda} - \frac{l}{2}\right) - \Phi\left(\frac{l}{\lambda} - \frac{l}{2}\right)\right\}\right\} + o(d^{-3/4}),
\end{aligned} \tag{4.8}$$

where $b_d = -\frac{d}{2}\log\left(\frac{1}{d}\sum_{i=1}^d (x_i^d)^2\right)$.

Proof. Note that

$$
\begin{aligned}
& \mathbb{E}_{\mathbf{Y}_1^{d-}}\left[1 \wedge \frac{\pi_d(\mathbf{Y}_1^d)}{\pi_d(\mathbf{x}^d)}\right] \\
= \; & \mathbb{E}\left[1_{\{-b_d+o(d^{-3/4})+Q_1^d+\sigma_d(x_1z_1+x_2z_2)\le 0\}}\left\{1 \wedge \exp\left(-\frac{1}{\lambda}Q_1^d - \frac{\sigma_d}{\lambda}(x_1z_1+x_2z_2)+o(d^{-3/4})\right)\right\}\right] \\
= \; & \mathbb{E}\left[1_{\{-b_d+o(d^{-3/4})+Q_1^d+\sigma_d(x_1z_1+x_2z_2)\le 0\}}\left\{1 \wedge \exp\left(-\frac{1}{\lambda}Q_1^d + o(d^{-3/4})\right)\right\}\right] \\
& -\frac{\sigma_d}{\lambda}(x_1z_1+x_2z_2)\mathbb{E}\left[1_{\{-b_d+o(d^{-3/4})+Q_1^d+\sigma_d(x_1z_1+x_2z_2)\le 0\}}\exp\left(-\frac{1}{\lambda}Q_1^d+o(d^{-3/4})\right); Q_1^d > 0\right] \\
& +o(d^{-3/4}). \qquad (4.9)
\end{aligned}
$$

The second equality follows by differentiating

$$
1 \wedge \exp\left(-\frac{1}{\lambda}Q_1^d - \frac{\sigma_d}{\lambda}(x_1z_1+x_2z_2)+o(d^{-3/4})\right)
$$

with respect to z_1 and z_2, see Breyer and Roberts (2000) page 192.

The lemma follows by applying Proposition 4.3 to (4.9). □

Let $V(\cdot) \in C_c^\infty$ be an arbitrary test function of the first two components only. For $b \ge 0$, $\lambda > 1$ and $u_1, u_2 \in \mathbb{R}$, let

$$
\begin{aligned}
& h_\lambda(b, u_1, u_2) \\
= \; & \frac{l^2}{2}\left\{\Phi\left(-\frac{l}{2}\right) + \exp\left(-\frac{l^2}{2}\times\frac{\mu_\lambda}{\lambda}\right)\left\{\Phi\left(\frac{b}{l}+\frac{l}{\lambda}-\frac{l}{2}\right) - \Phi\left(\frac{l}{\lambda}-\frac{l}{2}\right)\right\}\sum_{i=1}^{2}\frac{\partial^2}{\partial u_i^2}\right\}V(u_1,u_2) \\
- \; & l^2\sum_{i=1}^{2}u_i\frac{\partial}{\partial u_i}V(u_1,u_2)\exp\left(-\frac{l^2}{2}\times\frac{\mu_\lambda}{\lambda}\right)\left\{\frac{1}{l}\phi\left(\frac{b}{l}+\frac{l}{\lambda}-\frac{l}{2}\right)+\frac{1}{\lambda}\left\{\Phi\left(\frac{b}{l}+\frac{l}{\lambda}-\frac{l}{2}\right)-\Phi\left(\frac{l}{\lambda}-\frac{l}{2}\right)\right\}\right\}.
\end{aligned}
\qquad (4.10)
$$

Then Lemma 4.5 follows from Lemma 4.4. The proof is identical to Lemma 3.7, and so, the details are omitted.

Lemma 4.5 *For any $\lambda > 1$ and for $\mathbf{X}_0^d = \mathbf{w}^d \in F_d^{(\alpha,\gamma)}$,*

$$
\mathbb{E}\left[(V(\mathbf{Y}_1^d) - V(\mathbf{X}_0^d))\left\{1 \wedge \frac{\pi_d(\mathbf{Y}_1^d)}{\pi_d(\mathbf{X}_0^d)}\right\}\middle| \mathbf{X}_0^d = \mathbf{w}^d\right] = \frac{1}{d}h_\lambda(b_d, w_1, w_2) + o(d^{-5/4}).
$$

Corollary 4.6 follows immediately from Lemma 4.5 by straightforward but tedious integration, *c.f.* Lemma 3.7, Corollary 3.8 and Lemma 3.9.

Corollary 4.6 *For any* $\mathbf{w}^d \in F_d^{(\alpha,\gamma)}$ *and for any sequence of positive integers* $\{c_d\}$ *such that* $[d^\beta] \leq c_d \leq [d^\alpha]$*,*

$$d\mathbb{E}\left[(V(\mathbf{Y}^d_{c_d+1}) - V(\mathbf{X}^d_{c_d})) \left\{ 1 \wedge \frac{\pi_d(\mathbf{Y}^d_{c_d+1})}{\pi_d(\mathbf{X}^d_{c_d})} \right\} \middle| \mathbf{X}^d_0 = \mathbf{w}^d \right] \to \int_0^\infty h_\lambda(b, w_1, w_2)\mu_\lambda e^{-\mu_\lambda b}\, db, \tag{4.11}$$

where

$$\int_0^\infty h_\lambda(b, w_1, w_2)\mu_\lambda e^{-\mu_\lambda b}\, db = s(l) \sum_{i=1}^{2} \left\{ \frac{1}{2} \frac{\partial^2}{\partial w_i^2} V(w_i) - \frac{w_i}{2} \frac{\partial}{\partial w_i} V(w_i) \right\}. \tag{4.12}$$

Proof of Theorem 4.2. The theorem follows immediately from Corollary 4.6. The details of the proof are identical to the proof of Theorem 3.5.

5 Simulation Study

As noted in Section 4.1, it is difficult to prove results for general $f(\cdot)$. Therefore a simulation study was conducted to see to what extent (1.1) holds for general choices of $f(\cdot)$. Since Roberts *et al.* (1997) results only hold for continuous $f(\cdot)$, we restrict attention to continuous $f(\cdot)$. We follow Roberts and Rosenthal (1998) and Neal and Roberts (2006) in measuring the speed/efficiency of the algorithm in terms of the first-order efficiency. That is, for a multidimensional Markov chain $\mathbf{X}$ with first component X^1, say, the first-order efficiency is defined to be $\mathbb{E}[(X^1_{t+1} - X^1_t)^2]$, where $\mathbf{X}_t$ is assumed to be stationary. Throughout the simulation study we take $d = 50$ (similar results were obtained for $d = 20$ and $d = 100$) and all estimates are based on runs of $n = 250000$ iterations after a burn-in of 1000 iterations. We estimate $\mathbb{E}[(X^1_{t+1} - X^1_t)^2]$ by $\frac{1}{n}\sum_{i=1}^n (X^1_i - X^1_{i-1})^2$ and the acceptance rate is estimated by $\frac{1}{n}\sum_{i=1}^n 1_{\{\mathbf{X}_i \neq \mathbf{X}_{i-1}\}}$. We then plot acceptance rate against first-order efficiency.

A range of choices of $f(\cdot)$ were considered. The results presented are from four such choices of $f(\cdot)$ which are indicative of more general behaviour. The distributions considered are:-

$$f(x) \propto (1 + x^2/5)^3 \quad (x \in \mathbb{R}) \tag{5.1}$$

$$f(x) \propto \exp(-0.5(x - 0.5)^2) \quad (x \in \mathbb{R}) \tag{5.2}$$

$$f(x) \propto \exp(-0.5(x - 2)^2) \quad (x \in \mathbb{R}) \tag{5.3}$$

$$f(x) \propto \exp(-8(x - 0.5)^2) + \exp(-8(x + 0.5)^2) \quad (x \in \mathbb{R}) \tag{5.4}$$

The constrained Gaussian distributions of Section 4 have the properties that $f(\cdot)$ is unimodal and symmetric about its mean with mean 0. Of the chosen distributions, (5.1) is a standard t_5 distribution which has both these properties, (5.2) and (5.3) which are $N(0.5, 1)$ and $N(2, 1)$, respectively, are unimodal and symmetric about their means but have non-zero mean and (5.4) which is a mixture of $N(0.5, (0.25)^2)$ and $N(-0.5, (0.25)^2)$ is symmetric about its mean of 0 but is not unimodal. The results are presented in Figure 1 along with simulations from the uniform hypersphere (Section 3) for comparison.

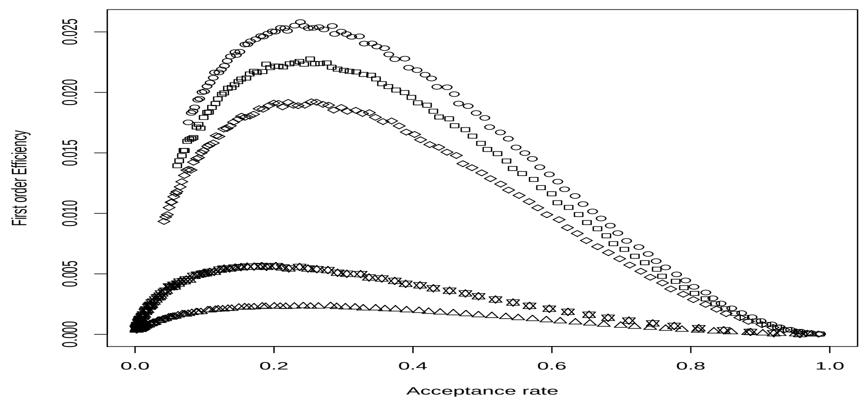

Figure 1: First-order efficiency $\mathbb{E}[(X_{t+1}^1 - X_t^1)^2]$, as a function of overall acceptance rates for:- a) Uniform Hypersphere, Section 3, (circles); b) t_5-distribution, (5.1), (squares); c) $N(0.5, 1)$, (5.2), (diamonds); d) $N(2, 1)$, (5.3), (stars); e) Mixture of Normals, (5.4), (triangles).

From Figure 1 it can be seen that the optimal acceptance rate for all the distributions except the $N(2, 1)$ is the same as for the uniform hypersphere. That is, the optimal acceptance rate is 0.234. For $N(2, 1)$ a lower acceptance rate was observed of approximately 0.182. This suggests that the results of Section 4 extend to densities, $f(\cdot)$ for which the modal value(s) lies between -1 and 1 but fails when the modal value(s) of the density $f(\cdot)$ lie outside this range. This was supported by a study of

$$f(x) \quad \propto \quad \exp(-0.5(x-2)^2) + \exp(-0.5(x+2)^2) \qquad (x \in \mathbb{R}), \tag{5.5}$$

a mixture of $N(-2, 1)$ and $N(2, 1)$ distributions, where the optimal acceptance rate was observed to be approximately 0.164.

Finally, the plots in figure 1 of all the distributions except (5.3) are very similar but on different scales. We define relative first-order efficiency as the first-order efficiency for l divided by the first-order efficiency for $\hat{l}$ where $\hat{l}$ is the optimal choice of l (*cf.* Corollary 3.6). Thus relative first-order efficiency takes values between 0 and 1 and represents the loss in efficiency from choosing suboptimal l. Therefore in Figure 2 we plot the estimated relative first-order efficiency against acceptance rate for the distributions plotted in Figure 1. The plots for all the distributions except (5.3) are indistinguishable. For the $N(2,1)$ and the mixture of Normals distribution, (5.5), the relative efficiency for an acceptance rate of 0.234 are approximately 0.96 and 0.94, respectively, suggesting that tuning the acceptance rate to 0.234 gives very good results in general.

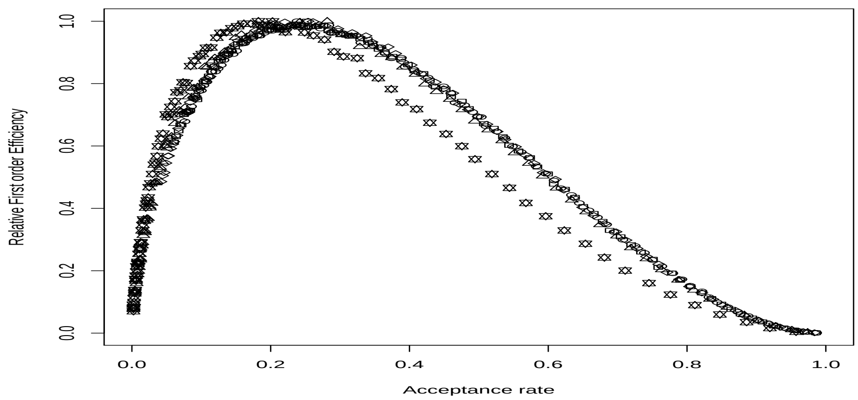

Figure 2: Relative first-order efficiency $d\mathbb{E}[(X_{t+1}^1 - X_t^1)^2]$, as a function of overall acceptance rates for:- a) Uniform Hypersphere, Section 3, (circles); b) t_5-distribution, (5.1), (squares); c) $N(0.5,1)$, (5.2), (diamonds); d) $N(2,1)$, (5.3), (stars); e) Mixture of Normals, (5.4), (triangles).

6 Summary

This paper has shown that the optimal scaling results of Roberts *et al.* (1997) extend to Gaussian distributions with a global (spherical) constraint. A simulation study has shown that the 0.234 rule is applicable more generally for a variety of distributions under the spherical constraint. However, as seen by densities (5.3) and (5.5) not all the results of Roberts *et al.* (1997) carry over to a spherical constraint. On the other hand, the tuning rule given by (1.1) still performs well.

The radial constraint is the key feature in these results. In particular, for the hypercube, Neal *et al.* (2007), rather different limiting results are observed. The major difference between the hypercube (the non-zero density is constrained to $\mathbf{x}^d \in [0,1]^d$) and the hypersphere (the non-zero density is constrained to $d^{-1}\sum_i x_i^2 \leq 1$), is that in the former case the discontinuity is local, depending upon individual components, whilst in the latter case the discontinuity is global, depending upon a function of all the components. In particular, the global constraint leads to continuous (Gaussian) limits for the distributions of the individual components.

In order to derive analytic results it has been necessary to restrict attention to constrained Gaussian distributions. For $\lambda > 1$ and $f(\cdot) \sim N(0, \lambda)$, the limiting behaviour of individual components are independent of λ. However, the limiting behaviour of the radial component is dependent upon λ. As previously mentioned, the case $\lambda < 1$ is not of great interest since the constraint is essentially redundant. For the case $\lambda = 1$, the statement of Theorem 4.2 holds but a very different proof is required.

Finally, the method of proof employed here can be used for other optimal scaling results where the acceptance probability is non-constant but is mixing at a much faster rate than the movement in individual components. This has been observed to be the case for the hypercube model studied in Neal *et al.* (2007) and for certain classes of non-IID target densities studied in Bédard (2006).

Acknowledgements

We would like to thank the referee for their valuable comments, in particular, the inclusion of a simulation study.

References

M. Bédard, On the Robustness of Optimal Scaling for Random Walk Metropolis Algorithms, Ph.D. Thesis, University of Toronto, 2006.

M. Bédard, ”Weak convergence of Metropolis algorithms for non-IID target distributions,” To appear in Ann. Appl. Prob., 2007.

J. Besag, and P.J. Green, ”Spatial statistics and Bayesian computation,” J. Roy. Statist. Soc. Ser. B 55 25–37, 1993.

J. Besag, P.J. Green, D. Higdon and K. Mengersen, ”Bayesian computation and stochastic systems,” Statist. Science 10 3–66, 1995.

P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.

L. Breyer, and G.O. Roberts, ”From Metropolis to diffusions: Gibbs states and optimal scaling,” Stochastic Process. Appl. 90 181–206, 2000.

S.N. Ethier, and T.G. Kurtz, Markov processes, characterization and convergence, Wiley, New York, 1986.

S.P. Meyn and R.L. Tweedie, Markov Chains and Stochastic Stability, Springer-Verlag, New York, 1993.

P.J. Neal and G.O. Roberts, "Optimal Scaling for partially updating MCMC algorithms," Ann. Appl. Prob. 16 475–515, 2006.

P.J. Neal, G.O. Roberts and W.K. Yuen, "Optimal Scaling of Random Walk Metropolis algorithms with Discontinuous target densities," submitted to Ann. Appl. Prob., 2007.

G.O. Roberts, A. Gelman and W.R. Gilks, "Weak convergence and optimal scaling of Random walk Metropolis algorithms," Ann. Appl. Prob. 7 110–120, 1997.

G.O. Roberts and J.S. Rosenthal "Optimal scaling of discrete approximations to Langevin diffusions," J. Roy. Statist. Soc. Ser. B, 60 255–268, 1998.

G.O. Roberts and J.S. Rosenthal, "Optimal scaling for various Metropolis-Hastings algorithms," Statist. Science 16 351–367, 2001.