Rate optimality of Random walk Metropolis algorithm in high-dimension with heavy-tailed target distribution

Kengo KAMATANI

Abstract

High-dimensional asymptotics of the random walk Metropolis-Hastings (RWM) algorithm is well understood for a class of light-tailed target distributions. We develop a study for heavy-tailed target distributions, such as the Student $t$-distribution or the stable distribution. The performance of the RWM algorithms heavily depends on the tail property of the target distribution. The expected squared jumping distance (ESJD) is a common measure of efficiency for light-tail case but it does not work for heavy-tail case since the ESJD is unbounded. For this reason, we use the rate of weak consistency as a measure of efficiency. When the number of dimension is $d$, we show that the rate for the RWM algorithm is $d^2$ for the heavy-tail case where it is $d$ for the light-tail case. Also, we show that the Gaussian RWM algorithm attains the optimal rate among all RWM algorithms. Thus no heavy-tail proposal distribution can improve the rate.

Keywords: Markov chain; Diffusion limit; Consistency; Monte Carlo; Stein’s method

1 Introduction

The Markov chain Monte Carlo (MCMC) method is a technique for the evaluation of the complicated integrals that uses a random sequence of Markov chains. In Bayesian community, it is used to calculate integrals with respect to the Bayesian posterior distribution. Although a lot of new techniques have been introduced for this calculation, MCMC is still one of the most popular tool for that purpose.

The random walk Metropolis (RWM) algorithm is the most widely used subclass of MCMC methods, due to its simplicity. It produces a simple modification of a symmetric random-walk Markov chain. It is applicable to both discrete state space and continuous state space. We focus on the latter and assume $\mathbb{R}^d$ as the state space in this paper.

Another explanation for the appeal of the RWM algorithm is its generality. Any kinds of symmetric random-walk can be used for this algorithm. We can use the Gaussian and the Student $t$-random walk or their mixture. Even if we decide to use the Gaussian random walk, we still have a free choice of its covariance structure. Note that although the underlying random-walk is not ergodic, the resulting Markov chain is usually ergodic (See Section 3.1 of Tierney [1994] for example).

However this generality means that, in turn, the practitioner should choose one particular symmetric random-walk Markov chain among all possibilities. Although any choice may produce ergodic Markov chain, the performance of the RWM algorithm depends on the random-walk. There are many practical strategies to monitor the performance of the MCMC, and it is advisable to use those for choosing an efficient random-walk for the RWM algorithm. There are also a lot of theoretical works for the performance of the MCMC. The detail will be described below.

Apart from the choice, the performance of the MCMC heavily depends on the target distribution, which is the underlying probability measure for the integral we want to approximate. For example, in Roberts and Rosenthal [1998], Metropolis-adjusted Langevin algorithm (MaLa) was shown to have better

*Supported in part by Grant-in-Aid for Young Scientists (B) 24740062.
rate of convergence than that of Gaussian RWM in high-dimensional unimodal target distribution. On the other hand if the target distribution is super-exponential, Gaussian RWM can be geometrically ergodic but MaLa cannot. Thus there is no uniformly optimal choice for the random-walk for RWM algorithm, and we should be careful for the structure of the target distribution.

A simple classification for the target distribution uses its tail property. If it is a light-tail distribution, geometric property of the RWM algorithm was studied in Mengersen and Tweedie 1996, Roberts and Tweedie 1996, and Jarner and Hansen 2000, by using Foster-Lyapunov type drift condition. It was proved in Mengersen and Tweedie 1996 and Jarner and Hansen 2000 that the exponential or lighter tail is necessary and sufficient for geometric ergodicity for the RWM algorithm. These asymptotic results correspond to the analysis for $M \to \infty$ where $M$ is the number of iteration of the MCMC. On the other hand, in Roberts et al. 1997, they considered asymptotic property of the RWM algorithm as the number of dimension $d \to \infty$. We will refer to this, high-dimensional asymptotics setting.

This seemingly curious setting brought useful insights for practitioners. Under this setting with the Gaussian random-walk, the asymptotic optimal choice of the covariance structure was obtained. Practically, this optimal choice can be obtained by monitoring simple statistic as described in Gelman et al. 1996. Another interesting results is that the convergence rate is on the order of $d$. This can be interpreted that the number of iteration $M$ should be proportional to $d$. This rate may vary among MCMC algorithms. For example, the MaLa has the order $d^{1/3}$, which is better than $d$ of the RWM algorithm Roberts and Rosenthal, 1998. The assumption in Roberts et al. 1997 is rather restrictive, and there has been a considerable effort for generalization of this result.

For heavy-tail target distributions, ergodic properties were studied in Jarner and Tweedie 2003, and Jarner and Roberts 2007. Their results show that the RWM algorithm can not have geometric ergodicity but polynomial ergodicity under heavy-tail target distribution. Also it was showed that a random-walk with heavy-tail increment distributions improves the rate of polynomial convergence. A natural question is whether this is true or not for high-dimensional setting. The convergence rate may or may not be worse than $d$, and this rate can or can not be improved by using heavy-tail increment distribution. To solve these questions is quite important in practical point of view.

Compared to ergodic property, there are only a few works for high-dimensional asymptotics for heavy-tail target distributions. One exception is Sherlock and Roberts 2009. They developed theoretical analysis on the expected squared jumping distance, which will be described later (see Section 2.2). Another related result is Neal and Roberts 2011, which is not for heavy-tail target distribution but for heavy-tail increment distribution of the random-walk of the RWM algorithm. However the above natural question is not solved yet. Main difficulty is that, unlike light-tail case, there are (at least) two kinds of convergence rate for heavy-tail target distribution. It makes the expected squared jump distance useless and optimality of the RWM algorithms becomes difficult to define.

In this paper, we shall concentrate on a class of heavy-tail target distributions under high-dimensional setting. We follow local consistency framework studied in Kamatani 2014 to overcome the difficulty mentioned above. Key result is that if we use the Gaussian random-walk in RWM algorithm, then the convergence rate is $d^2$, which is much worse than the light-tail case $d$. Surprisingly, this rate is optimal in a sense, and hence no heavy-tail increment distribution can improve the rate.

The paper is organized as follows. The assumptions and some background are given in Section 2. Since usual expected squared jumping distance does not make sense for heavy-tail target distributions, we define the rate of convergence as in Kamatani 2014. In Section 3, this notion will be tested to the light-tail, Gaussian target distribution, and then applied to rotationally symmetric heavy-tail distributions. Proofs are relegated to Section 4. In Section 5, we prepare a simple introduction for Stein’s technique, which is the main tool for our proofs.

We use the following notation throughout this paper. The state space is $\mathbb{R}^d$ throughout, and the Euclidean norm is denoted $\| \cdot \|$ and the inner product is denoted $\langle \cdot, \cdot \rangle$. Write $N_d(\mu, \Sigma)$ for $d$-dimensional normal distribution with the mean vector $\mu \in \mathbb{R}^d$ and the variance covariance matrix $\Sigma$, and $\phi_d(x; \mu, \Sigma)$ be its probability distribution function. The $d \times d$-identity matrix is denoted by $I_d$.

We also use the notation $\mathcal{L}(X)$ to denote the law of a random variable $X$. For $x \in \mathbb{R}$, write $x^+ = \max(x, 0)$.
max\{0, x\}, x^- = max\{0, -x\} and write \(|x|\) for the integer part of \(x \geq 0\). Write the sup norm by \(\|h\|_\infty = \sup_{z \in E} |h(z)|\) for \(h : E \to \mathbb{R}\) for a state space \(E\). If \(h\) is absolutely continuous, we write \(h'\) for the derivative.

2 Asymptotic properties of High-dimensional MCMC

2.1 Assumption for the target distribution

We consider a sequence of the target distributions \((P_d)_{d \in \mathbb{N}}\) indexed by the number of dimension \(d\). For a given \(d\), \(P_d\) is a \(d\)-dimensional probability distribution that is a scale mixture of the normal distribution. Furthermore, our asymptotic setting is that the number of dimension \(d\) goes infinity while the mixing distribution \(Q\) is unchanged.

Let \(Q(dy)\) be a probability measure on \([0, \infty)\). Let \(P_d\) be the scale mixture of the normal distribution defined by

\[
P_d = \mathcal{L}(X_0^d), \quad Q_d = \mathcal{L}(\|X_0^d\|^2/d)
\]

where \(X^d_0|Y \sim N_d(0, Y I_d)\) and \(Y \sim Q\). In particular, \(P_d\) is rotationally symmetric, that is, it is invariant under all orthogonal transform. We assume that the mixing distribution \(Q\) is \(\delta_1\) (the Dirac measure charging \(1 \in \mathbb{R}\)) or it satisfies the following, where we write \(f^{(i)}\) for the \(n\)-th derivative of a function \(f\).

**Assumption 1.** Probability distribution \(Q\) has the strictly positive probability distribution function \(q(y)\). The probability distribution function \(q(y)\) is two-times continuously differentiable and \(q^{(i)}(y)\) vanishes at \(+0\) and \(+\infty\) for \(i = 0, 1, 2\). Moreover, \(\lim_{y \to +\infty} yq(y) = 0\).

Let \(g(x; \alpha, \nu) \propto x^{\nu-1} \exp(-\alpha x)\) be the probability distribution function of the Gamma distribution \(G(\alpha, \nu)\). The following lemma says that the value of \(p_d(x)\) only depends on \(\|x\|^2\) and is a monotone decreasing function with respect to \(\|x\|\). This property is important for the random-walk Metropolis algorithm since it makes the acceptance probability function very simple.

**Lemma 2.1.** Under Assumption 1, both \(P_d\) and \(Q_d\) have the probability distribution functions \(p_d\) and \(q_d\) that satisfy

\[p_d(x) \propto \|x\|^{2-d} q_d(\|x\|/d).\]

Moreover, \(p_d(x_1) < p_d(x_2)\) if and only if \(\|x_1\| > \|x_2\|\).

**Proof.** Since \(P_d\) and \(Q_d\) are scale mixture of normal distribution and that of the chi-squared distribution with \(x\)-respectively, we have

\[p_d(x) = \int_0^\infty \phi_d(x; 0, z I_d)Q(dz), \quad q_d(y) = \int_0^\infty g(y; \frac{d}{2}z, \frac{d}{2})Q(dz)\]

and hence the first claim follows by the relation of integrands, that is, \(g(\|x\|/d; \frac{d}{2}, \frac{d}{2}) \propto \|x\|^{d-2}\phi_d(x; 0, z I_d)\).

The second claim comes from the fact that \(\phi_d(x_1; 0, z I_d) < \phi_d(x_2; 0, z I_d)\) if and only if \(\|x_1\| > \|x_2\|\). \(\square\)

The following lemma says that \(q_d\) is very close to \(q\) when \(d\) is sufficiently large. A proof is in Section C.

**Lemma 2.2 (Convergence of probability density function).** Under Assumption 1, \(\lim_{d \to \infty} \|q_d^{(i)} - q^{(i)}\|_\infty = 0\) for \(i = 0, 1, 2\).

**Example 1** (Student t-distribution). The probability distribution function of the \(t\)-distribution with \(\nu > 0\) degree of freedom is

\[p_d(x) = \frac{\Gamma((\nu + d)/2)}{\Gamma(\nu/2)^{d/2} \pi^{d/2}(1 + \|x\|^2/\nu)^{(\nu+d)/2}}\]

In this case, \(Q\) is the inverse chi-squared distribution with \(\nu\)-degree of freedom with probability distribution function \(q(y) \propto y^{\nu/2-1} e^{-y/(2\nu)}\). It is straightforward to check that \(q^{(i)}(y)\) vanishes at \(+0\) and \(+\infty\) for \(i = 0, 1, 2\) and \(\lim_{y \to +\infty} yq(y) = 0\). For properties of the (multivariate) \(t\)-distribution, see Kotz and Nadarajah [2004].
Example 2 (Stable distribution). If $P_d$ is the rotationally symmetric $\alpha$-stable distribution with characteristic function $\int \exp(-i(t,x))P_d(dx) = \exp(-\|t/2\|^{\alpha})$, then $Q$ is $\alpha/2$-stable distribution on the half line with Laplace transform $\int \exp(-ty)Q(dy) = \exp(-\|y\|^{\alpha/2})$. Although there is no closed form of probability density function $q(x)$, all derivatives of $q(x)$ are continuous and vanishes at 0 and $\infty$, and $q(x) \sim x^{-\alpha-1}$ as $x \to +\infty$. See Section 14 of Sato [1999].

Next we review some properties of the uniform distribution on the surface of the sphere $\{x \in \mathbb{R}^d; \|x\| = d\}$. Let $U^d = (U^d_i)_{i=1,\ldots,d}$ be uniformly distributed on the surface and $V^d = (V^d_i)_{i=1,\ldots,d}$ be an independent random variable on the surface. Then $(U^d, V^d) = \sum_{i=1}^d U^d_i V^d_i$ has the same law as $U^d$. The properties of $U^d_1$ are well-known. The law of $(U^d_1)^2/d$ is Beta$(1/2, (d-1)/2)$, and $U^d_1$ has the mean 0 and the variance 1. Asymptotic normality result with a sharp bound can be found in Diakonis and Freedman [1987] for the total variation distance and Chatterjee and Meckes [2008] for the Wasserstein distance.

Lemma 2.3. If $U^d = (U^d_i)_{i=1,\ldots,d}$ is uniformly distributed on the surface of the sphere $\{x \in \mathbb{R}^d; \|x\| = d\}$, then

$$\|\mathcal{L}(U^d_i) - N(0,1)\|_{TV} \leq \frac{8}{d-4}, \|\mathcal{L}(U^d_i) - N(0,1)\|_1 \leq \frac{3}{d-1}$$

where $\|P - Q\|_{TV} = 2 \sup_A |P(A) - Q(A)|$ and $\|P - Q\|_1 = \sup_{f \in B_1} |P(f) - Q(f)|$, where $B_1$ is the set of functions $f$ such that $\sup |f(x) - f(y)|/|x - y| \leq 1$.

Since $P_d$ is rotationally symmetric in this paper, above uniform bound is the key fact for our results. In addition to Lemma 2.3 we will use

$$\mathbb{E}[(U^d_i)^\alpha] = \frac{d^{\alpha/2}B(\frac{\alpha+1}{2}, \frac{d-1}{2})}{B(\frac{1}{2}, \frac{d-1}{2})} \to \frac{\Gamma(\frac{\alpha+1}{2})}{\Gamma(\frac{d}{2})} 2^{\alpha/2} \ (d \to \infty) \ (2.2)$$

for $\alpha > -1$, where we used Stirling’s approximation.

2.2 Expected squared jumping distance

We call that a probability measure $S_d$ on $\mathbb{R}^d$ is symmetric measure if $S_d(A) = S_d(-A)$ for any Borel set $A$, and we denote $S_d$ for the set of all symmetric probability measures in $\mathbb{R}^d$.

Let $P_d$ be a probability measure on $\mathbb{R}^d$ with density $p_d(x)$. In this paper, a Markov chain $X^d = (X^d_m)_{m \in \mathbb{N}}$ is called the random-walk Metropolis (RWM) chain for the target $P_d$ if $X^d_0$ is an $\mathbb{R}^d$-valued random variable, and for $m \geq 1$,

\[
\begin{align*}
Y^d_m &= X^d_{m-1} + W^d_m, \ W^d_m \sim S_d \\
X^d_m &= \begin{cases} 
Y^d_m & \text{with probability } \alpha_d(X^d_{m-1}, Y^d_m) \\
X^d_{m-1} & \text{with probability } 1 - \alpha_d(X^d_{m-1}, Y^d_m)
\end{cases}
\end{align*}
\]

where $\alpha_d(x,y) = \min\{1, p_d(y)/p_d(x)\}$. Write the law of $X^d$ by RWM($P_d, S_d$) if $X^d_0 \sim P_d$. It is well known that $X^d$ is a Markov chain reversible with respect to $P_d$, that is, if $X^d \sim \text{RWM}(P_d, S_d)$, then the random vectors

\[
(X^d_0, X^d_1, \ldots, X^d_m) \text{ and } (X^d_m, X^d_{m-1}, \ldots, X^d_0)
\]

have the same law.

Expected squared jumping distance (ESJD) is formally defined in Sherlock and Roberts [2009] as a measure of efficiency of MCMC as

$$\text{ESJD}(P_d, S_d) := \mathbb{E}[\|X^d - X^d_0\|^2].$$

The main result in Roberts et al. [1997] is that the (finite dimensional) probability law of $X^d$ converges weakly to the Langevin diffusion. The speed measure of the diffusion is proportional to $1/\text{ESJD}(P_d, S_d)$ in the limit, and hence ESJD can be interpreted as a measure of efficiency (larger is the better).
An interesting property for ESJD is the existence of optimality. Let $P_d$ satisfy (2.4) and recall that $Q_d$ is the law of $\|X_0^d\|^2/d$. Write $P_1$ for the cumulative probability distribution of $X_0^d$. Suppose that $t^2 P_1(-l_t/2)$ is bounded, and $l_t$ is its maximizer. Let $S_d \in S_d$ be a distribution on $\{x \in \mathbb{R}^d; \|x\| = l_t\}$. The following is more or less well-known.

**Proposition 2.1.** For each $d \in \mathbb{N}$, $\sup_{S_d \in S_d} \text{ESJD}(P_d, S_d) = \text{ESJD}(P_d, S_d^*) = 2(l^*)^2 P_1(-l^*/2)$.

**Proof.** Suppose that $X_0^d \sim P_d$ and $W_1^d \sim S_d$. Let

$$Z^d = \frac{\|X_0^d + W_1^d\|^2 - \|X_0^d\|^2}{2} = \langle X_0^d, W_1^d \rangle + \|W_1^d\|^2/2.$$

By Lemma 2.1, $p_d(X_0^d + W_1^d) < p_d(X_0^d)$ if and only if $\|X_0^d + W_1^d\| > \|X_0^d\|$, that is, $Z^d > 0$. Since the conditional laws of $Z^d - \|W_1^d\|^2$ and $-Z^d$ given $W_1^d$ are the same,

$$\mathbb{E}[p_d(X_0^d + W_1^d)/p_d(X_0^d)] \bigg| W_1^d > 0 = \int_{Z^d > 0} \frac{p_d(x + W_1^d)}{p_d(x)} p_d(x) dx = \int_{Z^d > 0} p_d(x + W_1^d) dx = \mathbb{P}(Z^d - \|W_1^d\|^2 > 0 | W_1^d) = \mathbb{P}(Z^d < 0 | W_1^d).$$

By using the above equation,

$$\mathbb{E}[1 \wedge \frac{p_d(X_0^d + W_1^d)}{p_d(X_0^d)} | W_1^d] = \mathbb{P}(Z^d \leq 0 | W_1^d) + \mathbb{E}\left[\frac{p_d(X_0^d + W_1^d)}{p_d(X_0^d)}, Z^d > 0 | W_1^d\right]$$

$$= 2 \mathbb{P}(Z^d \leq 0 | W_1^d) = 2 P_1(-\|W_1^d\|^2/2)$$

since $(X_0^d, W_1^d / \|W_1^d\|) \sim P_1$. Thus by definition,

$$\text{ESJD}(P_d, S_d) = \mathbb{E}[1 \wedge \frac{p_d(X_0^d + W_1^d)}{p_d(X_0^d)} \|W_1^d\|^2]$$

$$= 2 \mathbb{E}[\|W_1^d\|^2 P_1(-\|W_1^d\|^2/2)] \leq 2(l^*)^2 P_1(-l^*/2).$$

The properties of ESJD and expected acceptance ratio have been studied since the seminal work of Roberts et al., [1997]. However for heavy-tail case, $t^2 P_1(-l_t/2)$ is not bounded in general, and hence it can not be used as a measure of efficiency. Thus for heavy-tail target distribution, we need another measure of efficiency, and we use the framework of Kamatani, [2014].

### 2.3 Consistency for high dimensional MCMC

In this section, we review consistency of MCMC studied in Kamatani, [2014]. Set a sequence of Markov chains $\{\xi^d := (\xi^d_m; m \in \mathbb{N}_0)\} \ (d \in \mathbb{N})$ with the invariant probability measures $\{\Pi_d\}_d$. The sequence $\{\xi^d\}_d$ (or its law) is called consistent if

$$\frac{1}{M} \sum_{m=0}^{M-1} f(\xi^d_m) - \Pi_d(f) = o_P(1) \quad (2.3)$$

for any $M, d \to \infty$ for any bounded continuous function $f$. This says that the integral $\Pi_d(f)$ we want to calculate is approximated by Monte Carlo simulated value $\frac{1}{M} \sum_{m=0}^{M-1} f(\xi^d_m)$ after a reasonable number of iteration $M$. Regular Gibbs sampler should satisfy this type of property (more precisely, local consistency. See Kamatani, [2014]) when $d$ is the sample size of the data. However this is not always the case as described in the end of Kamatani, [2014]. In our case, (2.3) is not satisfied in two respects: the state space is not the same for $d \in \mathbb{N}$, and $M = M_d$ should satisfy a certain rate.
Recall that the target distribution in the current study is rotationally symmetric, and hence there are two sufficient statistics:

$$\frac{X_d^d}{\|X_0^d\|} \text{ and } \|X_0^d\|^2$$  \hspace{1cm} (2.4)

where $X_0^d \sim P_d$. For $d = 1, 2, \ldots$, let $X^d = (X^d_m, m \in \mathbb{N}_0)$ be an $\mathbb{R}^d$-valued stationary process with invariant distribution $P_d$ on $(\Omega, \mathcal{F}, \mathbb{P})$.

We consider asymptotic properties of $X^d$ through $d \to \infty$. As commented above, the state space for $X^d$ ($d \in \mathbb{N}$) changes as $d \to \infty$ that is inconvenient for further analysis. To overcome the difficulty, we set a projection $\pi_E = \pi_{d,E}$ for a finite subset $E \subset \{1, \ldots, d\}$ by

$$\pi_E(x) = (x_i)_{i \in E} \ (x = (x_i)_{i=1,\ldots,d}).$$

We denote $\pi_k$ for $\pi_{\{1,\ldots,k\}}$. We also define $r_d : \mathbb{R}^d \to \mathbb{R}$ to be $\sqrt{d(\|x\|^2/d - 1)}$ or $\|x\|^2/d$ depending on the target distribution $P_d$. Roughly, the properties after projections $\pi_E$ ($E \in \{1, \ldots, d\}$) and $r_d$ correspond to that of $X^d_m/\|X^d_m\|$ and $\|X^d_m\|$ with respectively.

**Definition 1** (Consistency). We call that a sequence of $\mathbb{R}^d$-valued Markov chain $\{X^d\}_{d \in \mathbb{N}}$ and its law are consistent if

$$\frac{1}{M_d} \sum_{m=0}^{M_d-1} f \circ \pi_{E_d^d}(X^d_m) - P_d(f \circ \pi_{E_d^d}) = o_P(1)$$  \hspace{1cm} (2.5)

as $d \to \infty$ for any $k \in \mathbb{N}$, $M_d \to \infty$ and for any bounded continuous function $f : \mathbb{R}^k \to \mathbb{R}$ and any $k$-elements $E_d^k$ of $\{1, \ldots, d\}$. Also, for $r_d : \mathbb{R}^d \to \mathbb{R}$, we call $X^d$ and its law are consistent with respect to $r_d$ if

$$\frac{1}{M_d} \sum_{m=0}^{M_d-1} f \circ r_d(X^d_m) - P_d(f \circ r_d) = o_P(1)$$  \hspace{1cm} (2.6)

for any $M_d \to \infty$ and for any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$.

Note that consistency (2.5) does not imply consistency with respect to $r_d$ (2.6). However, these properties are, in fact, prepared just for explanation of weak consistency below, and this definition itself is impractical in our current setting; for high-dimensional case ($d \to \infty$), consistency (with or without projection) is rarely satisfied. As in Kamatani [2013], we relax the condition for $M_d$ and introduce the rate of consistency, which is the key of our results.

**Definition 2** (Weak Consistency). We call that a sequence of $\mathbb{R}^d$-valued Markov chain $\{X^d\}_{d \in \mathbb{N}}$ and its law are weakly consistent with rate $T_d$ if (2.5) is satisfied for any $M_d \to \infty$ such that $M_d/T_d \to \infty$. If (2.6) is satisfied in place of (2.5) we call that $\{X^d\}$ and its law are weakly consistent with rate $T_d$ with respect to $r_d$.

In this paper, we compare different MCMC methods by the rate of weak consistency. This is just a formalization of the usual approach in this area; this type of comparison for high-dimensional MCMC is at least dates back to Roberts and Rosenthal [1998]. An important remark to note here is that the weak consistency corresponds to two sufficient statistics (2.4) may vary. For our heavy-tail setting, the rate for weak consistency for $X_0^d \mapsto X_0^d/\|X_0^d\|$ and $X_0^d \mapsto \|X_0^d\|^2/d$ are $d$ and $d^2$ with respectively (see Propositions 3.3 and 3.4).

There is a technical difficulty for the analysis ($\pi_{E_d^k}(X^d_m))_{m \in \mathbb{N}}$ since it is not a Markov chain. However, if we pair the sequence with $r_d(X^d_m)$ then it become a Markov chain. The proof is easy and is omitted for brevity.

**Lemma 2.4.** If $P_d$ is rotationally symmetric, and $S_d$ is a symmetric measure, then both $(r_d(X^d_m))_{m \in \mathbb{N}_0}$ and $(\pi_{E_d^k}(X^d_m), r_d(X^d_m))_{m \in \mathbb{N}_0}$ are Markov chains if $X^d \sim \text{RWM}(P_d, S_d)$.
3 Main results

3.1 Weak consistency for Gaussian target distribution

First, we consider asymptotic property of \( r_d(X_m^d) \). For Gaussian case \( P_d = N_d(0, I_d) \), we set \( r_d : \mathbb{R}^d \to \mathbb{R} \) as

\[
    r_d(x) = \sqrt{d\left(\frac{\|x\|^2}{d} - 1\right)} \quad (x \in \mathbb{R}^d).
\]

The law of \( r_d(X_0^d) \), which is the stationary distribution of \( Y^{d,1} \) defined below, tends to \( N(0,2) \). This normal distribution is also the stationary distribution of the limit Ornstein-Uhlenbeck process. Let \([x] \) be the integer part of \( x \geq 0 \), and let \( \mu_k(\sigma) = \mathbb{E}[|Z|^k \exp(-Z^2)] \) for \( Z \sim N(\sigma^2/2, \sigma^2) \).

**Proposition 3.1.** \( P_d = N_d(0, I_d) \). Set \( X^d \sim \text{RWM}(P_d, S_d) \) for \( S_d = N_d(0, \frac{I_d^2}{d} I_d) \), and let \( Y_t^{d,1} = r_d(X_{[dt]}^d) \). Then \( Y^{d,1} \) converges to \( Y^1 \) that is the solution of

\[
    dY_t^1 = -\frac{\sigma(l)^2}{4} Y_t^1 dt + \sigma(l) dW_t; Y_0^1 \sim N(0,2) \tag{3.1}
\]

where \( \sigma(l)^2 = 4\mu_2(l) \) and \((W_t)_{t \geq 0}\) is the standard Wiener process. In particular, \( \text{RWM}(P_d, S_d) \) has weak consistency with rate \( T_d = d \) with respect to \( r_d \).

As suggested by Lemma 2.3, the Gaussian proposal distribution attains the optimal rate for weak consistency with respect to \( r_d \). Therefore no heavier-tail proposal distribution can be better.

**Theorem 3.1.** \( P_d = N_d(0, I_d) \). For \( \text{RWM}(P_d, S_d) \) for \( S_d \in \mathcal{S}_d \), the optimal rate for weak consistency with respect to \( r_d \) is \( T_d = d \). In particular, \( S_d = N_d(0, \frac{I_d^2}{d} I_d) \) attains the optimal rate.

Next we consider asymptotic properties of \( \pi_{E^k_d}(X_m^d) \). If \( S_d \) is rotationally symmetric, then the law of \( \{\pi_{E^k_d}(X_m^d)\}_m \) does not depend on any choice of the \( k \)-elements \( E^k_d \) from \{1, \ldots, d\}. Therefore, to prove (2.5), it is sufficient to show for \( E^k_d = \{1, \ldots, k\} \) and in the following, we only consider \( \pi_k = \pi_{\{1, \ldots, k\}} \). Note that in the following, \( Y^2 \) does not mean the square of \( Y \).

**Proposition 3.2.** Let \( P_d = N_d(0, I_d) \). Set \( X^d \sim \text{RWM}(P_d, S_d) \) for \( S_d = N_d(0, \frac{I_d^2}{d} I_d) \), and let \( Y_t^d = (r_d(X_{[dt]}^d), \pi_k(X_{[dt]}^d)) \). Then \( Y^d \) converges to \((Y^1, Y^2)\) where \( Y^1 \) is the solution of (3.1) and \( Y^2 \) is the solution of

\[
    dY_t^2 = -\frac{\sigma(l)^2}{2} Y_t^2 dt + \sigma(l) dB_t^k; Y_0^2 \sim N_k(0, I_k) \tag{3.2}
\]

where \( \sigma(l)^2 = l^2 \mu_0(l) \) and \((B_t^k)_{t \geq 0}\) is the \( k \)-dimensional standard Wiener process independent of \((W_t)_{t \geq 0}\).

As in Theorem 3.1, this rate is optimal for weak consistency.

**Theorem 3.2.** For \( P_d = N_d(0, I_d) \) and for \( S_d = N_d(0, \frac{I_d^2}{d} I_d) \), \( \text{RWM}(P_d, S_d) \) is weakly consistent with rate \( T_d = d \). This rate is optimal.

3.2 Weak consistency for heavy-tailed target distribution

Now we focus on heavy-tailed distribution, that is, \( Q \) has a probability density \( q \). In this case, unlike the Gaussian case, the projection \( \sqrt{d(\|x\|^2/d - 1)} \) is not appropriate for the study of the trajectories of MCMC, since the law of \( \sqrt{d(\|X_m^d\|^2/d - 1)} \) as \( d \to \infty \) is not tight. Therefore we consider another projection

\[
    r_d(x) = \frac{\|x\|^2}{d}.
\]

By this projection, the law of \( r_d(X_m^d) \) tends to \( Q \) as \( d \to \infty \).
Proposition 3.3. Let \( Q \) satisfy Assumption \([1]\). Set \( X^d \sim \text{RWM}(P_d, S_d) \) for \( S_d = N_d(0, \frac{2}{d} I_d) \), and let \( Y^{d,1}_t = r_d(X^d_{[d,t]}) \). Then \( Y^{d,1} \) converges to a stationary ergodic process \( Y^1 \) that is the solution of
\[
\mathrm{d}Y^1_t = a(Y^1_t) \mathrm{d}t + \sqrt{b(Y^1_t)} \mathrm{d}W_t; Y^1_0 \sim Q
\]
where
\[
a(y) = 2(y + (\log y)'(y) y^2) \mu_2(l/\sqrt{y}) + y l^2 \mu_1(l/\sqrt{y}), \quad b(y) = 4 y^2 \mu_2(l/\sqrt{y})
\]
and \((W_t)_{t \geq 0}\) is the standard Wiener process. In particular, \( \text{RWM}(P_d, S_d) \) has weak consistency with rate \( T_d = d^2 \) for the projection \( r_d \).

As in the case of the Gaussian target distribution, the above rate is optimal for weak consistency with respect to \( r_d \). This is rather counter intuitive, since in practice and in ergodic theory, sometimes heavier proposal works well. The interpretation will be discussed in Section 3.3.

Theorem 3.3. Let \( Q \) satisfy Assumption \([1]\). For \( \text{RWM}(P_d, S_d) \) for \( S_d \in S_d \), the rate for weak consistency with projection \( r_d \) can not exceed \( d^{2-\epsilon} \), that is, \( r_d/d^{2-\epsilon} \to \infty \) for any \( \epsilon > 0 \). In particular, \( S_d = N_d(0, \frac{2}{d} I_d) \) attains the optimal rate \( T_d = d^2 \) in this sense.

Next we show asymptotic properties of \( \pi_E(X^d_m) \). The key fact is that the convergence rate for \( \|X^d_m\| \) and \( X^d_m/\|X^d_m\| \) are different. This makes analysis little bit complicated. However, thanks to Lemma \([3,3]\), we can prove weak consistency even for this case.

Proposition 3.4. Let \( Q \) satisfy Assumption \([1]\). Set \( X^d \sim \text{RWM}(P_d, S_d) \) for \( S_d = N_d(0, \frac{2}{d} I_d) \), and let \( Y^d_{k} = \langle r_d(X^d_{[d,k]}), \pi_k(X^d_{[d,k]}) \rangle \). Then \( Y^d \) converges to \( (y, Y^2(y)) \) where \( y \sim Q \) and \( Y^2 = Y^2(y) \) is the solution of
\[
\mathrm{d}Y^2_t = -\frac{\sigma(l)^2}{2} Y^2_t \mathrm{d}t + y^{1/2} \sigma(l) \mathrm{d}B^k_t; Y^2_0 \sim N_k(0, y I_k)
\]
where \( \sigma(l)^2 = l^2 \mu_0(l), \quad I^2 = l^2/y \) and \((B^k_t)_{t \geq 0}\) is the \( k \)-dimensional standard Wiener process.

Note that the process \( Y^2 \) is stationary and ergodic conditioned on \( y \). However it is not ergodic unconditional on \( y \). To obtain an ergodic result, we should consider the time scaling \( t \to d^2 t \). Moreover, this time scaling is the optimal rate, as in Theorem \([3,3]\).

Theorem 3.4. Let \( Q \) satisfy Assumption \([1]\). For \( S_d = N_d(0, \frac{2}{d} I_d) \), \( \text{RWM}(P_d, S_d) \) is weakly consistent with rate \( T_d = d^2 \). Moreover, the rate for weak consistency can not exceed \( d^{2-\epsilon} \) for any \( \epsilon > 0 \) and hence \( \text{RWM}(P_d, S_d) \) attains the optimal rate in this sense.

3.3 Discussion

- As studied in Theorems \([3,3]\) and \([3,3]\) the convergence rate for the heavy-tail target distribution is \( d^2 \), which is much worse than that for the light-tail target distribution. Though the rate is quite bad, it is optimal as described in those theorems.

- The optimality of Gaussian random walk Metropolis algorithm is rather counter intuitive in two respects. In \( \text{Jarner and Roberts} \ [2007] \), they showed that heavy-tail increment distribution improves the polynomial rate of convergence. It may seem a contradiction, but it is not. High-dimensional asymptotics corresponds to the local property of the MCMC but the ergodic theory corresponds to the global property. Thus the conclusion of Theorem \([3,4]\) and \( \text{Jarner and Roberts} \ [2007] \) are for two different properties. If we choose the initial state \( X^d_0 \) properly (See Lemma 4 of \( \text{Kamatani} \ [2014] \) and Section B of \( \text{Kamatani} \ [2012] \)), then local property may be more useful for the choice of \( S_d \). On the other hand if we have no idea for the choice of \( X^d_0 \), global property may provide more information for convergence.

8
In practice, it is also discussed the usefulness of heavy-tail increment distribution for heavy-tail target distribution. In our setting, the target distribution belongs to a class of unimodal rotationally symmetric distributions. If the target distribution is multi-modal, then it may be better to use heavy-tail increment distribution, though the theoretical validation for this might be difficult. If the target distribution is considered to be a perturbation of a rotationally symmetric distribution, then our conclusion may be applicable, and in that case, the Gaussian random-walk Metropolis algorithm attains the optimal rate.

• Only the optimal rate was discussed. For light-tail case, optimality of the limit process was also considered in the literature. It may also be possible for our case for the choice of \( l \) when the increment distribution is \( S_d = N_d(0, \frac{l^2}{d} I_d) \). Optimal choice of \( S_d \) for all \( S_d \) in this sense is even more difficult, since the limit can be a diffusion, a pure-jump process and a diffusion with jumps. Further work should be done for this.

• For a light-tail target distribution, expected squared jumping distance and expected acceptance rate work as good measures of efficiency in practice. For heavy-tail case, it is not easy to find such a statistic. One possibility is to estimate

\[
E[\log \| X^d_0 \| - \log \| X^d_0 \|]
\]

that is finite even for heavy-tail case and is on the order of \( d^{-1} \). If this value is moderately large, the MCMC may have a good performance. If the target distribution is far from rotationally symmetric, I recommend to normalise (scale properly) the distribution in advance.

• There is no known good strategy for heavy-tail target distribution that improves the convergence rate \( d^2 \). MaLa behaves like RWM as discussed in [Jarner and Roberts 2007] and [Kamatani 2009], and RWM with heavy-tail proposal does not improve the convergence rate. If the target is a perturbation of a stable distribution, then a discretization of the stable related process may work well. However this is out of the scope of this paper.

4 Proofs

Some notation is required for the proofs.

Random variables Sections 4.1-4.3 deal with the Gaussian target distribution \( P_d = N_d(0, I_d) \). In this case, we analyze the properties of the random variable

\[
Z^d := \langle X^d_0, W^d_1 \rangle + \frac{\| W^d_1 \|^2}{2}, \quad X^d_0 \sim P_d, W^d_1 \sim S_d.
\]

(4.1)

Note that \( \| X^d_0 + W^d_1 \|^2 - \| X^d_0 \|^2 = 2Z^d \), and the acceptance probability \( \alpha_d(X^d_0, X^d_0 + W^d_1) \) is

\[
\alpha(Z^d) := \exp(-Z^d+)
\]

for the Gaussian target distribution, where we write \( Z^d+: = \max\{Z^d, 0\} \). Moreover, the conditional distribution of \( Z^d \) given \( W^d_1 \) is \( N(\| W^d_1 \|^2/2, \| W^d_1 \|^2) \).

Sections 4.4-4.6 are for heavy-tail target distribution. Write \( y = r_d(X^d_0) = \| X^d_0 \|^2/d \). In this case, we analyze the random variable

\[
Z^d := \frac{\| X^d_0 + W^d_1 \|^2 - \| X^d_0 \|^2}{2\| X^d_0 \|^2/d} = \langle \frac{d^{1/2}X^d_0, d^{1/2}W^d_1}{\| X^d_0 \|^2}, \frac{d^{1/2}W^d_1}{\| X^d_0 \|^2} \rangle + \frac{1}{2} \frac{d^{1/2}W^d_1}{\| X^d_0 \|^2}
\]

\[
= \langle X^d_0, W^d_1 \rangle + \frac{1}{2}\| W^d_1 \|^2.
\]
Note that \(\|X_0^d + W_1^d\|^2 - \|X_0^d\|^2 = 2Z^d y\). Also, by Lemma 2.1 the acceptance probability \(\alpha_d(X_0^d, X_0^d + W_1^d)\) is

\[
\alpha_d(Z^d; y) := \min\left\{ \frac{p_d(X_0^d + W_1^d)}{p_d(X_0^d)}, 1 \right\}
\]

\[
= \min\left\{ \left( \frac{\|X_0^d + W_1^d\|}{\|X_0^d\|} \right)^{-d} q_d(\|X_0^d + W_1^d\|^2/d), 1 \right\}
\]

\[
= (1 + \frac{2Z^d}{d})^{-d} q_d(y + \frac{2Z^d y}{d})/q_d(y).
\]

For both Gaussian and heavy-tail target distributions, we will use

\[
U_1^d = \sqrt{d}\left( \frac{X_0^d}{\|X_0^d\|}, \frac{W_1^d}{\|W_1^d\|} \right).
\]

In both cases, the law of this random variable is the first element of the uniform distribution on the surface of the sphere \(\{x \in \mathbb{R}^d; \|x\| = d\}\) conditional on \(W_1^d\).

**Probability measures** In the following six sections, we write

\[
\mathbb{P}_y(\cdot) = \mathbb{P}(\cdot | r_d(X_0^d) = y),
\]

\[
\mathbb{P}_{y,\sigma}(\cdot) = \mathbb{P}(\cdot | r_d(X_0^d) = y, \|W_1^d\| = \sigma),
\]

\[
(4.2)
\]

\[
\mathbb{P}_{x,y}(\cdot) = \mathbb{P}(\cdot | \pi_k(X_0^d) = x, r_d(X_0^d) = y).
\]

Denote \(N_\sigma\) for \(N(\sigma^2/2, \sigma^2)\). Some important properties of this probability distribution \(N_\sigma\) are described in Section \(\Delta\).

**Constants** Let \(W_{1,k}^d\) be the \(k\)-th element of the vector \(W_1^d \in \mathbb{R}^d\). We will use notation

\[
y = r_d(X_0^d), \quad \sigma = \|W_1^d\|, \quad \sigma_k = |W_{1,k}^d|.
\]

We have \(\sigma^2 = \sum \sigma_k^2\).

We use the following result of triangular arrays that is a slight modification of Lemma 9 of Genon-Catalot and Jacod [1993]. The proof follows the same lines as the proof of their lemma via Lenglart inequality, and is omitted. For \(d \in \mathbb{N}\), let \(\{G_i^d\}_i\) be a filtration and let \(\chi_{i,\alpha}^d\) be \(G_i^d\)-measurable random variables indexed by \(i \in \mathbb{N}_0\) and \(\alpha \in \mathcal{A}_d\). We denote \(X_{i,\alpha}^d \to 0\) uniformly if the convergence \(\mathbb{P}(|X_{i,\alpha}^d| > \epsilon) \to 0\) \((d \to \infty)\) holds uniformly in \(\alpha \in \mathcal{A}_d\) for any \(\epsilon > 0\). Let \(T_d\) be a \(\mathbb{N}_0\)-valued sequence.

**Lemma 4.1.** The following two conditions imply \(\sup_{i < T_d} \left| \sum_{j=0}^{i} \chi_{j,\alpha}^d \right| \mathbb{P} \to 0 \) uniformly:

\[
\sum_{i < T_d} |\mathbb{E}[\chi_{i,\alpha}^d | G_{i-1}^d]| \mathbb{P} \to 0 \text{ uniformly,} \quad (4.3)
\]

\[
\sum_{i < T_d} \mathbb{E}[|\chi_{i,\alpha}^d|^2 | G_{i-1}^d] \mathbb{P} \to 0 \text{ uniformly.} \quad (4.4)
\]

To highlight the dependence on \(S_d\), we may write \(\mathbb{P}_{S_d}\) for the underlying probability measure. For a sequence of random variables \(V^d\), we denote \(V^d \mathbb{P} \to 0 \) uniformly if the convergence \(\mathbb{P}_{S_d}(|V^d| > \epsilon) \to 0\) is uniformly in \(S_d \in S_d \) for any \(\epsilon > 0\).
4.1 Proof of Proposition 3.1

Write
\[ \sigma_d^2 = \| W_d^2 d \|_2^2 = \sigma^2 \left( 1 + \frac{y}{d} \right). \] (4.5)

By definition, \( Z^d = \sigma_d U_1 + \sigma^2 / 2 \).

**Lemma 4.2.** Set \( d \geq 2 \), \( \sigma > 0 \) and \( y \in \mathbb{R} \), so that \( \sigma_d^2 > 0 \). For any absolutely continuous \( N_{\sigma} \)-integrable function \( h \),
\[ |E_{y,\sigma}[h(Z^d) - N_{\sigma}h + d^{-1/2} \sigma^2 y N_{\sigma}f']| \leq d^{-1} \left( 6 + 3yd^{-1/2} + 4y^2 \sqrt{2/\pi} \right) \sigma \| h' \|_\infty \]
where \( f \) is the solution of Stein’s equation (A.3). In particular, if \( h(z) = ze^{-z^2} \),
\[ |E_{y,\sigma}[h(Z^d) + d^{-1/2} \frac{h_2(\sigma)}{2}]| \leq d^{-1} \left( 6 + 3yd^{-1/2} + 4y^2 \sqrt{2/\pi} \right) \sigma. \]

**Proof.** Set \( \tilde{Z}^d = \sigma_d U_1 + \sigma^2 / 2 \) and \( U_1 \sim N(0, 1) \). By Lemma 4.2,
\[ |E_{y,\sigma}[h(Z^d)]| \leq \frac{3\sigma_d}{d-1} \| h' \|_\infty \leq \frac{6\sigma_d}{d} \| h' \|_\infty \] (4.6)
if \( d - 1 \geq d/2 \). If \( f = \tilde{g}(h, N_{\sigma}) \) is the solution of (A.3),
\[ E_{y,\sigma}h(\tilde{Z}^d) - N_{\sigma}h = \sigma^2 E_{y,\sigma}[f'(\tilde{Z}^d)] - E_{y,\sigma}[\tilde{Z}^d - \frac{\sigma^2}{2}]f(\tilde{Z}^d) \]
\[ = \sigma^2 E_{y,\sigma}[f'(\tilde{Z}^d)] - \sigma d E_{y,\sigma}[U_1 f(\tilde{Z}^d)] \]
\[ = (\sigma^2 - \sigma_d^2) E_{y,\sigma}[f'(\tilde{Z}^d)] \] (4.7)
where the last equation comes from Stein's characteristic (A.2). Now let \( g = \tilde{g}(f', N_{\sigma}) \). Then applying the above equation to \( f' \) in place of \( h \) we have
\[ |E_{y,\sigma}f'(\tilde{Z}^d) - N_{\sigma}f'| \leq |\sigma^2 - \sigma_d^2||g'||_\infty. \]
Furthermore, by (A.6), we have \( ||g'||_\infty \leq 4\sigma^{-2}||f'||_\infty \leq 4\sigma^{-3} \sqrt{2/\pi} ||h'||_\infty \) and hence
\[ |E_{y,\sigma}f'(\tilde{Z}^d) - N_{\sigma}f'| \leq |\sigma^2 - \sigma_d^2|4\sigma^{-3} \sqrt{2/\pi} ||h'||_\infty. \]
Therefore, by this inequality together with (4.6) and (4.7), we have
\[ |E_{y,\sigma}[h(Z^d) - N_{\sigma}h + (\sigma^2 - \sigma_d^2) N_{\sigma}f']| \leq d^{-1} \left( 6\sigma_d + d(\sigma^2 - \sigma_d^2)4\sigma^{-3} \sqrt{2/\pi} \right) ||h'||_\infty. \]
By putting \( \sigma^2 - \sigma_d^2 = -d^{-1/2} \sigma^2 y \) and \( \sigma_d \leq \sigma(1 + y/2\sqrt{d}) \), the former inequality in the lemma follows. Finally, the last claim follows by \( N_{\sigma}h = 0, ||h'||_\infty \leq 1 \) and by Lemma A.2 when \( h(z) = ze^{-z^2} \).

**Proof of Proposition 3.1** For \( \epsilon < 1 \), let \( K(\epsilon) = [\epsilon, \epsilon^{-1}] \) be a closed interval in \((0, \infty)\), and let \( f \) be the solution of (A.3) for \( h(z) = ze^{-z^2} \) with \( \sigma = ||W_1^1|| \). Set
\[ \begin{align*}
  a_d(y) & := dE_{y}[r_d(X_1^d) - r_d(X_0^d)], \\
  b_d(y) & := dE_{y}[\{r_d(X_1^d) - r_d(X_0^d)\}^2], \\
  c_d(y) & := dE_{y}[\{r_d(X_1^d) - r_d(X_0^d)\}^4].
\end{align*} \] (4.8)
By Lemma 4.2,
\[ a_d(y) = 2\sqrt{d}E_{y}[h(Z^d)] = -yE_y[\mu_2(\sigma)] + O(d^{-1/2}) = -y\mu_2(\ell) + o(1) \]
uniformly in \( y \in K(\epsilon) \) by the dominated convergence theorem. For diffusion part, we have

\[
    b'_d(y) = 4E_y[(Z^d)^2 e^{-Z^d}] =: 4E_y[h_2(Z^d)].
\]

Conditional distribution of \( Z \) given \( y \in K(\epsilon) \) converges to \( N_t \), and \( h_2(Z^d) \) is uniformly integrable by \( E_y[(U)^4] = 3d/(d+2) \) (by (2.2)) and \( E_y[\sigma_d^2] = t^4(1+2d^{-1}) \). Thus

\[
    b_d(y) = 4N_t h_2 + o(1) = 4\mu_2(l) + o(1)
\]

uniformly in \( y \in K(\epsilon) \). Finally we check

\[
    c_d(y) = d^{-1/2}4E_y[(Z^d)^4 e^{-Z^d}] \leq d^{-1/2}4E_y[(Z^d)^4] = O(d^{-1}).
\]

Thus the first claim follows by the convergence of Markov chain to a diffusion process (Theorem IX 4.21 of Jacod and Shiryaev (2003)). The last claim is a conclusion from Lemma B.1.

\[\]

4.2 Proof of Theorem 3.1

For two random variables \( X \) and \( Y \), we write \( X \prec Y \) if there exists \( C > 0 \) such that \( \mathbb{P}_{S_d}(|X| > \epsilon) \leq C \mathbb{P}_{S_d}(|Y| > \epsilon) \) for any \( \epsilon > 0 \) and \( S_d \in S_d \). In the proof of Theorem 3.1, we use the fact

\[
    \|X^d_i\|^2 - \|X^d_0\|^2 \prec \langle X^d, \frac{W^d}{\|W^d\|} \rangle^2.
\]

Note that the random variable \( \langle X^d, \frac{W^d}{\|W^d\|} \rangle \) follows the standard normal distribution, which is, of course, free from the choice of \( S_d \). This can be proved by reversibility;

\[
    \mathbb{P}_{S_d}(\|X^d_i\|^2 - \|X^d_0\|^2 > \epsilon) = 2\mathbb{P}_{S_d}(\|X^d_i\|^2 - \|X^d_0\|^2 < -\epsilon) \leq 2\mathbb{P}_{S_d}(\|X^d_0 + W^d\|^2 - \|X^d_0\|^2 < -\epsilon) = 2\mathbb{P}_{S_d}(\|W^d\|^2 + 2\|W^d\|\langle X^d_0, \frac{W^d}{\|W^d\|} \rangle < -\epsilon) = 2\mathbb{P}_{S_d}(\|W^d\|^2 + \langle X^d_0, \frac{W^d}{\|W^d\|} \rangle^2 - \langle X^d_0, \frac{W^d}{\|W^d\|} \rangle^2 < -\epsilon) \leq 2\mathbb{P}_{S_d}(\langle X^d_0, \frac{W^d}{\|W^d\|} \rangle^2 < -\epsilon).
\]

**Lemma 4.3.** For \( T_d/d \to 0 \)

\[
    \sup_{i < j < T_d} |r_d(X^d_i) - r_d(X^d_j)| \xrightarrow{p_s} 0 \text{ uniformly.}
\]

**Proof.** Set \( \sigma_d \) as in (4.5) and set

\[
    \chi^d_k = \begin{cases} 
    r_d(X^d_k) - r_d(X^d_{k-1}), & \text{if } k < T_d, \\
    \chi^d_k \mathbb{1}_{\{||W_0^d||^2 < d\}}, & \text{otherwise.}
    \end{cases}
\]

By the definition of (4.1),

\[
    \chi^d_1 = \begin{cases} 
    2d^{-1/2} Z^d, & \text{with probability } e^{-Z^d}, \\
    0, & \text{otherwise.}
    \end{cases}
\]

Furthermore, we set

\[
    A_d = \sup_{i < j < T_d} \left| \sum_{k=i+1}^{j} \chi^d_k \right|, B_d = \sum_{0 \leq k < T_d} \left| E_{S_d}[\chi^d_k f_{k-1}] \right|, C_d = \sum_{0 \leq k < T_d} \mathbb{E}_{S_d}\left[ \left\{ \chi^d_k \right\}^2 \right| f_{k-1}
\]

12
and also
\[ \hat{A}_d = \sup_{i < j < T_d} \left| \sum_{k=i+1}^j \chi_k^d \right|, \quad \hat{B}_d = \sum_{0 \leq k < T_d} \left| E_{S_d} \left[ \chi_k^d \left| F_{k-1}^d \right. \right] \right|, \quad \hat{C}_d = \sum_{0 \leq k < T_d} E_{S_d} \left[ \{ \chi_k^d \}^2 \right] F_{k-1}^d \]
where the filtration is defined by
\[ F_k^d = \sigma \{ \| X_0^d \|^2, \ldots, \| X_k^d \|^2, \| W_1^d \|^2, \ldots, \| W_{k+1}^d \|^2 \} \quad (k \in \mathbb{N}_0). \]  
(4.11)

Note that \( E_{S_d}[|F_0^d|] = E_{y,\sigma}[\cdot] \), and \( \hat{B}_d \leq B_d \) and \( \hat{C}_d \leq C_d \) by \( F_{k-1}^d \)-measurability of \( \| W_k^d \| \). Our claim is \( A_d \overset{p}{\rightarrow} 0 \) uniformly. Thanks to reversibility, this is equivalent to
\[ \sup_{S_d} \mathbb{P}_{S_d} \left( \inf_{i < j < T_d} \sum_{k=i+1}^j \chi_k^d \leq -\epsilon \right) = o(1). \]

By dividing \( \chi_k^d = \chi_k^d 1\{\| W_k^d \|^2 \geq d \} + \tilde{\chi}_k^d \), the left-hand side is bounded above by
\[ \sup_{S_d} \mathbb{P}_{S_d} \left( \inf_{i < T_d} \sum_{k=i+1}^j \chi_k^d 1\{\| W_k^d \|^2 \geq d \} < 0 \right) + \sup_{S_d} \mathbb{P}_{S_d} \left( \hat{A}_d \geq \epsilon \right). \]  
(4.12)

We will show that the both two terms tend to 0. Note here that \( E_{y,\sigma}[(U_1^d)^2] = 1 \) and \( E_{y,\sigma}[y] = 0 \). By stationarity, the first term in the above is bounded by
\[ \sup_{S_d} T_d \mathbb{P}_{S_d}(\chi_1^d < 0, \| W_1^d \| \geq d^{1/2}) \leq \sup_{S_d} T_d \mathbb{P}_{S_d}(Z^d < 0, \| W_1^d \| \geq d^{1/2}) \]
\[ \leq \sup_{S_d} T_d \mathbb{P}_{S_d}(U_1^d < -\frac{\sigma^2}{2\sigma_d}, \sigma \geq d^{1/2}) \]
\[ \leq T_d \mathbb{P}(U_1^d < -\frac{d^{1/2}}{2(1+yd^{-1/2})^{1/2}}) \]
\[ \leq T_d \mathbb{E} \left[ 4d^{-1}(1 + yd^{-1/2})(U_1^d)^2 \right] = 4Td^{-1} = o(1) \]
where we used Schwartz inequality in the fourth inequality. Next we prove \( \hat{A}_d \overset{p}{\rightarrow} 0 \). By Lemma 4.1, it is sufficient to show \( \hat{B}_d \overset{p}{\rightarrow} 0 \) and \( \hat{C}_d \overset{p}{\rightarrow} 0 \) uniformly. Now we show the convergence of \( \hat{B}_d \). By stationarity,
\[ \mathbb{P}_{S_d}(\hat{B}_d > \epsilon) \leq \epsilon^{-1} \mathbb{E}_{S_d}[\hat{B}_d] \leq \epsilon^{-1} T_d \mathbb{E}_{S_d} \left[ E_{y,\sigma}[\chi_1^d], \sigma < d^{1/2} \right]. \]

It is sufficient to show the boundedness of \( \sup_{S_d} d \mathbb{E}_{S_d} \left[ E_{y,\sigma}[^{\sigma}d(0)_1] \right], \sigma < d^{1/2} \). By taking \( h(z) = ze^{-z} \), direct application of Lemma 4.2 together with the fact that \( \sup_{\sigma} \mu_2(\sigma) < \infty \) yields
\[ \sup_{S_d} d \mathbb{E}_{S_d} \left[ E_{y,\sigma}[^{\sigma}d(0)_1] \right], \sigma < d^{1/2} \]
\[ \leq 2 \sup_{S_d, \sigma < d^{1/2}} d^{1/2} \mathbb{E}_{S_d} \left[ (d^{-1/2}|y|\frac{\mu_2(\sigma)}{2}) + d^{-1}(6 + 3yd^{-1/2} + 4y^2\sqrt{2/\pi}) \sigma \right] < \infty. \]

Thus the convergence of \( \hat{B}_d \) follows. Finally we show \( \hat{C}_d \overset{p}{\rightarrow} 0 \). This is sufficient to show \( C_d \overset{p}{\rightarrow} 0 \) uniformly, and this is also sufficient to show the boundedness of \( \sup_{S_d} d \mathbb{E}_{S_d} \left[ |\chi_1^d|^2 \right] \) by stationarity. However by 4.9 for some constant \( c > 0 \),
\[ d \mathbb{E}_{S_d}[|\chi_1^d|^2] \leq c \mathbb{E}_{S_d}[|X_0^d - W_1^d||W_1^d|^2] = 3c \]
since \( (X_0^d, \frac{W_1^d}{\| W_1^d \|}) \sim N(0, 1) \). Thus \( C_d \overset{p}{\rightarrow} 0 \) uniformly, and hence the claim follows. \[ \square \]
Proof of Theorem 4.1. Suppose now that we can take a rate $T_d \to \infty$ such that it is better than the rate $d$, that is, $T_d/d \to 0$. We can find a number of iteration $M_d \to \infty$ so that $M_d/T_d \to \infty$ and $M_d/d \to 0$. Then by Lemma 4.3, we have

$$\sup_{m \leq M_d} |r_d(X^d_m) - r_d(X^d_m)| = o_p(1),$$

and hence

$$M_d^{-1} \sum_{m=0}^{M_d-1} f \circ r_d(X^d_m) - P_d(f \circ r_d) = (f \circ r_d)(X^d_0) - P_d(f \circ r_d) + o_p(1)$$

where $Y^d_0 \sim N(0, 2)$. Therefore this value can not be $o_p(1)$ unless $f$ is a constant. Thus $X^d$ is not weak consistent with the rate $T_d$ and hence the claim holds.

4.3 Proof of 3.2 and Theorem 3.2

Proof of Proposition 3.2. Set

$$\begin{pmatrix} a_d(x, y) := d\mathbb{E}_{x,y}[\pi_k(X^d_1) - \pi_k(X^d_0)], \\ b_d(x, y) := d\mathbb{E}_{x,y}[\pi_k(X^d_0) - \pi_k(X^d_0)^{\otimes 2}], \\ c_d(x, y) := d\mathbb{E}_{x,y}[\pi_k(X^d_0) - \pi_k(X^d_0)^{\otimes 4}] \end{pmatrix}$$ (4.13)

By Stein’s identity (A.1),

$$a_d(x) = d\mathbb{E}_{x,y}[\pi_k(W^d_1)^d)\alpha(Z^d)]$$

Conditioned on $X^d_0$, the random variable $\langle X^d_0, W^d_1 \rangle$ follows $N(0, 1 + y/\sqrt{d})$. Therefore, for $Z \sim N_1$, by the bounded convergence theorem, we have

$$\mathbb{E}_{x,y}[\alpha'(Z^d)] = \mathbb{E}[\alpha'(Z)] + o(1) = -\frac{1}{2} \mu_0(l) + o(1)$$

uniformly for $y \in K(\varepsilon) = [\varepsilon, \varepsilon^{-1}]$ for any $\varepsilon < 1$, where we used (A.7) and Lemma A.1 to show $\mathbb{E}[\alpha'(Z)] = -\mu_0(l)/2$. For the diffusion term, observe that

$$b_d(x, y) = d\mathbb{E}_{x,y}[\pi_k(W^d_1)^\otimes 2 \alpha(Z^d)]$$

By Stein’s identity (A.2),

$$d\mathbb{E}_{x,y}[(\pi_k(W^d_1)^d)\otimes 2 - \frac{l^2}{d} I_k)\alpha(Z^d)] = l^2\mathbb{E}_{x,y}[\pi_k(X^d_0 + W^d_1) \otimes \pi_k(W^d_1) \alpha'(Z^d)] = o(1).$$

By the dominated convergence theorem, we have

$$b_d(x, y) = l^2 I_k \mathbb{E}_{x,y}[\alpha(Z^d)] + o(1) = l^2 \mu_0(l) I_k + o(1)$$

uniformly for $y \in K(\varepsilon)$. Finally,

$$c_d(x, y) = d\mathbb{E}_{x,y}[\pi_k(W^d_1)^4] = o(1).$$

Therefore by the convergences of (1.8) and (4.13), the step Markov process $(r_d(X^d_{\lfloor t \rfloor}), \pi_k(X^d_{\lfloor t \rfloor}))_{t \geq 0}$ converges to the joint process of $Y^1$ and $Y^2$ defined in Propositions 3.1 and 3.2 by Theorem IX 4.21 of Jacod and Shiryaev 2003.

□

14
Proof of Theorem 3.2. The first claim is a direct conclusion from Lemma 3.1 and Proposition 3.2. We show optimality. To show this, it is sufficient to show that for any $T_d \to \infty$ such that $T_d/d \to \infty$, there exists $k_d \in \{1, \ldots, d\}$ such that

$$\sup_{i < s_d} |X_{i,k_d}^d - X_{0,k_d}^d| \xrightarrow{p} 0$$

uniformly as $d \to \infty$ where $X_{i,k_d}^d$ is the $k$-th element of the vector $X_i^d \in \mathbb{R}^d$. By Lemma 4.1 together with stationarity, it is sufficient to show

$$\sum_{k=1}^d |\mathbb{E}_{S_d}[|X_{1,k}^d - X_{0,k}^d|]| = O(1) \quad (4.14)$$

and

$$\sum_{k=1}^d \mathbb{E}_{S_d}[|X_{1,k}^d - X_{0,k}^d|^2] = O(1) \quad (4.15)$$

uniformly in $S_d \in S_d$. Indeed, we can choose $k = k_d \in \{1, \ldots, d\}$ so that $\mathbb{E}_{S_d}[|X_{1,k}^d - X_{0,k}^d|] + \mathbb{E}_{S_d}[|X_{1,k}^d - X_{0,k}^d|^2]$ is the smallest among all $k \in \{1, \ldots, d\}$, and for that $k_d$, we can apply Lemma 4.1 to $\{X_{i,k_d}^d\}_i$.

The left-hand side of (4.15) is bounded above by the ESJD, and hence bounded as proved in Proposition 2.2. We show (4.14). Let $W_{1,k}^d$ be the $k$-th element of the vector $W_1^d$. Write $\sigma_k = |W_{1,k}^d|$ and $\sigma^2 = \sum_{k=1}^d \sigma_k^2 = \|W_1^d\|^2$. Then

$$\mathbb{E}_{S_d}[X_{1,k}^d - X_{0,k}^d|X_{0,k}^d, \sigma_k] = \mathbb{E}_{S_d}[W_{1,k}^d \exp(-Z_{d^+})|X_{0,k}^d, \sigma_k]$$

$$= \sigma_k \mathbb{E}_{S_d}[\frac{W_{1,k}^d}{\|W_1^d\|} \exp(-Z_{d^+})|X_{0,k}^d, \sigma_k].$$

Since $S_d$ is a symmetric measure, $W_{1,k}^d/\|W_1^d\|$ takes +1 and −1 for the same probability $p_d \in [0, 1/2]$. Thus the above becomes

$$\sigma_k p_d \mathbb{E}_{S_d}[\exp(-Z_{d^+}) - \exp(-\tilde{Z}_{d^+})|X_{0,k}^d, \sigma_k],$$

where $\tilde{Z}^d = Z^d - 2X_{0,k}^d W_{1,k}^d$. Conditioned on $W_1^d$, the law of $Z^d$ and $\tilde{Z}^d$ are $N(\sigma^2/2, \sigma^2)$ with the correlation $\rho = 1 - 2\sigma_k^2/\sigma^2$. In particular, $Z^d$ and $\tilde{Z}^d$ are exchangeable, that is, the law of $(Z^d, \tilde{Z}^d)$ and $(\tilde{Z}^d, Z^d)$ are the same. Thus by exchangeability, we have

$$\mathbb{E}_{S_d}[|X_{1,k}^d - X_{0,k}^d|] = \mathbb{E}_{S_d}[|\sigma_k p_d \{\exp(-Z_{d^+}) - \exp(-\tilde{Z}_{d^+})\}|X_{0,k}^d|] \leq 2^{-1} \mathbb{E}_{S_d}[\sigma_k \{\exp(-Z_{d^+}) - \exp(-\tilde{Z}_{d^+})\}]$$

$$= \mathbb{E}_{S_d}[\sigma_k \{\exp(-Z_{d^+}) - \exp(-\tilde{Z}_{d^+})\} 1\{Z_{d^+} > \tilde{Z}_{d^+}\}],$$

$$\mathbb{E}_{S_d}[\sigma_k \exp(-Z_{d^+}) - \exp(-\tilde{Z}_{d^+}) 1\{Z_{d^+} > \tilde{Z}_{d^+}\}]$$

$$= \mathbb{E}_{S_d}[\sigma_k \exp(-Z_{d^+}) - \exp(-\tilde{Z}_{d^+})(1 - 21\{Z_{d^+} > \tilde{Z}_{d^+}\})].$$

The conditional distribution of $\tilde{Z}^d$ given $Z^d$ and $W_1^d$ is $N(\sigma^2/2 + \rho(Z^d - \sigma^2/2), \sigma^2(1 - \rho^2))$ and hence

$$\mathbb{P}(Z^d > \tilde{Z}^d|Z^d, W_1^d) = \Phi\left(\frac{(1 - \rho)(Z^d - \sigma^2)}{\sigma \sqrt{1 - \rho^2}}\right).$$

Thus by simple calculation, we have

$$\left|\Phi(0) - \mathbb{P}(Z^d > \tilde{Z}^d|Z^d, W_1^d)\right| \leq \Phi(\infty) \left|\frac{(1 - \rho)(Z^d - \sigma^2)}{\sigma \sqrt{1 - \rho^2}}\right|$$

$$\leq \Phi(\infty) \left|\frac{(1 - \rho)(Z^d - \sigma^2)}{\sigma \sqrt{1 - \rho^2}}\right| = \sqrt{2}\Phi(\infty) \left|\frac{\sigma_k \sigma^2}{\sigma \sqrt{1 - \rho}}\right| = \sqrt{2}\Phi(\infty) \left|\frac{\sigma_k \sigma^2}{\sigma \sqrt{1 - \rho}}\right|.$$
Therefore, the right-hand side of (4.16) is

\[
2\mathbb{E}_{S_d} \left[ \sigma_k \exp(-Z^{d+}) (\Phi(0) - \Phi(Z^d > Z^d|Z^d)) \right] \leq 2\sqrt{2} \|\Phi\|_\infty \mathbb{E}_{S_d} \left[ \frac{\sigma_k^2}{\sigma^2} |Z^d - \frac{\sigma^2}{2}| \exp(-Z^{d+}) \right].
\]

Therefore the left-hand side of (4.14) is bounded above by

\[
2\sqrt{2} \|\Phi\|_\infty \mathbb{E}_{S_d} \left[ |Z^d - \frac{\sigma^2}{2}| \exp(-Z^{d+}) \right] \leq 2\sqrt{2} \|\Phi\|_\infty \sup_{\sigma > 0} (\mu_1(\sigma) + \frac{\sigma^2}{2}\mu_0(\sigma)) < \infty.
\]

Thus the claim follows.

\[\square\]

### 4.4 Proof of Proposition 3.3

We will denote \(I^2 = l^2/y\). If \(W_1^d \sim N_d(0, \frac{l^2}{d} I_d)\) then \(W_1^d = y^{-1/2}W_1^d \sim N_d(0, \frac{1}{d} I_d)\) conditional on \(\|X_0^d\|\).

**Lemma 4.4.** For \(W_1^d \sim N_d(0, \frac{l^2}{d} I_d)\) and for any absolutely continuous \(N_d\)-integrable function \(h\),

\[
|\mathbb{E}_y h(Z^d)| - N_1 h + \frac{3l^4}{2d} N_1 f'' + \frac{l^4}{2d} N_1 f' \leq \frac{C(l)}{d^2} \|h'\|_\infty
\]

where \(f = \mathcal{F}(h, N_1)\) and \(C(l)\) is a polynomial of \(l\).

**Proof.** By using Stein’s solution \(f\) together with (A.2), we have

\[
\mathbb{E}_y [h(Z^d)] - N_1 h = \mathbb{E}_y [f'(Z^d)] - \mathbb{E}_y [(Z^d - \frac{l^2}{2}) f(Z^d)]
\]

\[
= \mathbb{E}_y [f'(Z^d)] - \mathbb{E}_y [\langle \tilde{X}_0^d, W_1^d \rangle f(Z^d)] - \frac{l^2}{d} \mathbb{E}_y [(\|W_1^d\|^2 - I^2) f(Z^d)]
\]

\[
= \mathbb{E}_y [f'(Z^d)] - \frac{l^2}{d} \mathbb{E}_y [\langle \tilde{X}_0^d, W_1^d \rangle f(Z^d)]
\]

\[
= -\frac{3l^2}{2d} \mathbb{E}_y [\langle X_0^d, W_1^d \rangle f'(Z^d)] - \frac{l^2}{2d} \mathbb{E}_y [(\|W_1^d\|^2 - I^2) f'(Z^d)]
\]

(4.17)

where we used \(\|X_0^d\|^2 = d\). By this, together with the estimate \(\|f'\|_\infty \leq 4l^{-2} \|h'\|_\infty\) by (A.6), we obtain

\[
|\mathbb{E}_y [h(Z^d)] - N_1 h| \leq \frac{3l^3}{2d} + \frac{l^4}{2d} \|f'\|_\infty \leq \frac{C_0(l)}{d} \|h'\|_\infty
\]

(4.18)

where \(C_0(l) = 6l + 2l^2\) since \(\mathbb{E}_y [\langle X_0^d, W_1^d \rangle] \leq \mathbb{E}_y [\langle X_0^d, W_1^d \rangle^2]^{1/2} = 1\) and \(\mathbb{E}_y [\|W_1^d\|^2] = l^2\). Furthermore, applying (A.1) to (4.17), we obtain

\[
\mathbb{E}_y [h(Z^d)] - N_1 h = -\frac{3l^2}{2d} \mathbb{E}_y [\langle X_0^d, W_1^d \rangle f'(Z^d)] - \frac{l^2}{2d} \mathbb{E}_y [(\|W_1^d\|^2 - I^2) f'(Z^d)]
\]

(4.19)

\[
= -\frac{3l^4}{2d} \mathbb{E}_y [\langle X_0^d, W_1^d \rangle f''(Z^d)]
\]

\[
- \frac{l^2}{2d^2} \mathbb{E}_y [(\|W_1^d\|^2 - I^2) f''(Z^d)]
\]

\[
= -\frac{3l^4}{2d} \mathbb{E}_y [f''(Z^d)] + \frac{l^4}{2d} \mathbb{E}_y [f'(Z^d)]
\]

\[
- \frac{1}{2d} \mathbb{E}_y [(\|W_1^d\|^2 - I^2) f''(Z^d)]
\]

\[
= -A_d(y) - B_d(y).
\]
Then applying (4.18) to \( f'' \) and \( f' \) in place of \( h \), together with (A.6),

\[
|A_d - \{ \frac{3l^4}{2d} N_l f'' + \frac{l^4}{2d} N_l f' \} | \leq \frac{C_0(l)}{d} (\frac{3l^4}{2d} ||f''||_{\infty} + \frac{l^4}{2d} ||f'||_{\infty} ) \leq \frac{C_1(l)}{d^2} ||h'||_{\infty}
\]

where \( C_1(l) = C_0(l)(3l^2 + l^4/\sqrt{2}\pi) \). Finally, we get

\[
|B_d| \leq (\frac{2l^5}{d^2} + \frac{l^6}{2d^2}) ||f''||_{\infty} \leq \frac{C_2(l)}{d^2} ||h'||_{\infty}
\]

by (A.6) were \( C_2(l) = 4l^3 + l^4 \). Thus the claim holds for \( C(l) = C_1(l) + C_2(l) \).

\[\square\]

For \( \epsilon < 1 \), let \( K(\epsilon) = [\epsilon, \epsilon^{-1}] \) be a closed interval in \((0, \infty)\). Recall the definition of \( \alpha_d(z; y) \) in the beginning of Section 4. See Section C.2 for the proof of Lemma 4.5.

**Lemma 4.5.** Let \( h_d(z; y) = z\alpha_d(z; y) \) and \( h(z) = ze^{-z^+} \). Set

\[
\begin{align*}
M^d(\epsilon) &= d \sup_{y \in K(\epsilon)} ||h_d(; y) - h||_{\infty}, \\
M^d(\epsilon) &= d \sup_{y \in K(\epsilon)} ||\partial_z h_d(; y) - h'||_{\infty}.
\end{align*}
\]

Under Assumption 1, for any \( \epsilon < 1 \), \( \limsup_{d \to \infty} M^d(\epsilon) < \infty \) and \( \limsup_{d \to \infty} M^d(\epsilon) < \infty \).

**Proof of Proposition 3.3.** Set

\[
\begin{align*}
a_d(y) &:= dE_y[r_d(X_1^f) - r_d(X_0^d)], \\
b_d(y) &:= dE_y[(r_d(X_1^f) - r_d(X_0^d))^2], \\
c_d(y) &:= dE_y[r_d(X_1^f) - r_d(X_0^d)^4].
\end{align*}
\]

Then

\[
a_d(y) = 2dyE_y[h_d(Z^d; y)].
\]

For a positive number \( y \), let \( f_d(; y) = \mathfrak{F}(h_d(; y), N_1) \). Then by Lemma 4.5 we can apply Lemma 4.4 and obtain

\[
|a_d(y) - 2dyN_l h_d(; y) + 3y^4N_l (\partial_2^2 f_d(; y)) + y^4N_l (\partial_z f_d(; y))| \\
\leq \frac{2\epsilon^{-1}C(1)}{d} ||\partial_z h_d(; y)||_{\infty} = O(d^{-1})
\]

uniformly in \( y \in K(\epsilon) \). Suppose now that \( f = \mathfrak{F}(h, N_1) \) for \( h(z) = ze^{-z^+} \). Then by the property of Stein’s solution (A.6),

\[
\begin{align*}
\|\partial_2^2 f_d(; y) - f''\|_{\infty} &\leq 2l^{-2}\|\partial_z h_d(; y) - h'\|_{\infty} \leq 2l^{-2}d^{-1} M^d(\epsilon) = O(d^{-1}), \\
\|\partial_z f_d(; y) - f'\|_{\infty} &\leq 4l^{-2}\|\partial h_d(; y) - h\|_{\infty} \leq 4l^{-2}d^{-1} M^d(\epsilon) = O(d^{-1}).
\end{align*}
\]

(4.21)

By Taylor’s expansion and by Lemma 2.7 for \( z > 0 \),

\[
\lim_{d \to \infty} d(h_d(z; y) - h(z)) = \lim_{d \to \infty} d\left\{ z(1 + \frac{2z}{d})^{\frac{2z}{d}} \left( \frac{q_d(y + \frac{2z}{d})}{q_d(y)} - 1 \right) \\
+ z\left( (1 + \frac{2z}{d})^{\frac{2z}{d}} - (1 + \frac{2z}{d})^2 \right) + z\left( (1 + \frac{2z}{d})^{-\frac{2z}{d}} - e^{-z} \right) \right\} \\
= 2z^2 ye^{-z}(\log q)'(y) + 2z^2 e^{-z} + z^3 e^{-z}.
\]

17
Note that the left-hand side is 0 if \( z \leq 0 \). Using this, together with Lemma A.3, by the dominated convergence theorem, we have

\[
dN_1h_d(\cdot:y) = dN_1(h_d(\cdot:y) - h) \\
= 2y(\log q)'(y)\mathbb{E}[(Z^+)^2e^{-Z^+}] + 2\mathbb{E}[(Z^+)^2e^{-Z^+}] + \mathbb{E}[(Z^+)^2e^{-Z^+}] + o(1) \\
= (1 + y(\log q)'(y))\mu_2(1) + \frac{1}{2}\mu_3(1) + o(1)
\]

(4.22)

uniformly in \( y \in K(\epsilon) \) where \( Z \sim N_1 \). Therefore, by applying Lemmas A.1 and A.2 together with (4.20) and (4.22), we have

\[
a_d(y) = 2dyN_1h_d(\cdot:y) - 3y^4N_1(\partial_z f_d(\cdot:y)) - y^4N_1(\partial_z f_d(\cdot:y)) + o(1) \\
= 2y\{1 + y(\log q)'(y))\mu_2(1) + \frac{1}{2}\mu_3(1)\} - 3y^4N_1f'' - y^4N_1f' + o(1) \\
= 2(y + y^2(\log q)'(y))\mu_2(1) + y(-\frac{1}{2}\mu_2(1) + 2\mu_2^2(1)) + y^2(\mu_1(1) - \mu_2(1)) \\
\]

Also

\[
b_d(y) = 4y^2\mathbb{E}_y[(Z^d)^2\alpha_d(Z^d; y)] - 4y^2\mu_2(1) + o(1).
\]

For the convergence of \( b_d \), by the uniform integrability of \( (Z^d)^2 \) and by \( \alpha_d(z; y) \to a(z) \), we have

\[
b_d(y) = 4y^2\mathbb{E}_y[(Z^d)^2\alpha_d(Z^d; y)] = 4y^2\mu_2(1) + o(1).
\]

Also

\[
c_d(y) = d^{-2}8y^4\mathbb{E}_y[(Z^d)^4\alpha_d(Z^d; y)] = o(1).
\]

Therefore the convergence to the process \( Y^1 \) follows by Theorem IX 4.21 of Jacod and Shiryaev [2003]. Now we check ergodic property for the solution of the limit stochastic differential equation. We check that \( Q(dx) \) is the stationary distribution of the process. To see this, we show

\[
\frac{(\log q)'b + b'}{2} = a
\]

(4.23)

and

\[
\lim_{y \to \infty} q(y)b(y) = 0.
\]

(4.24)

For these calculation, we should recall the definition \( \Gamma^2 = \ell^2/y \). First, we note that

\[
\frac{(\log q)'b}{2} = 2(\log q)'(y)y^2\mu_2(1).
\]

and hence it is sufficient for (4.23) to show

\[
b'/2 = 2y\mu_2(1) + y\ell^2\mu_1(1).
\]

However this is clear by Lemma A.1 since

\[
b'/2 = 4y\mu_2(1) + 2y^2\mu_2^2(1)\ell \]

\[
= 4y\mu_2(1) - y\ell^2(\mu_2(1) - \mu_1(1)) = 2y\mu_2(1) + y\ell^2\mu_1(1).
\]

For the proof of (4.21), we split \( q(y)b(y) \) into \( yq(y) \) times \( y^{-1}b(y) \). By Assumption II the former term tends to 0, and the latter term is

\[
y^{-1}b(y) = 4y\mu_2(1) = 4y(-\frac{1}{2}\mu_1(1) + \ell^2\mu_0(1)) \leq 4y\ell^2\mu_0(1) = 4\ell^2\mu_0(1) \leq 4\ell^2\sup_{\sigma}\mu_0(\sigma) < \infty
\]
and hence (4.24) follows. Then by Theorem 2.3 of [Bibby et al. 2005], the limit process is ergodic. Indeed, thanks to (4.23), the conditions in Theorem 2.3 can be checked by
\[
\int_0^x a(y)q(y)dy = \int_0^x \left( \frac{(\log y)b + b'}{2} \right) q(y)dy = \int_0^x \frac{q'b + b'q}{2} dy = q(x)b(x)/2
\]
with the fact that \( qb \) vanishes at \(+0\) and \(+\infty\) by (4.24). Finally we apply Proposition B.1 for weak consistency.

\[\square\]

### 4.5 Proof of Theorem 3.3

**Lemma 4.6.** Set \( T_d = d^{2-\alpha} \) for some \( \alpha \in (0, 2) \). Then
\[
\sup_{i<j<T_d} |r_d(X_i^d) - r_d(X_j^d)| \overset{P}{\to} 0 \text{ uniformly.}
\]

**Proof.** We can certainly assume that \( \alpha \in (0, 1/2) \). Fix \( \beta \in (0, \alpha) \). By reversibility and tightness of \( y := r_d(X_0^d) \), it is sufficient to show
\[
\sup_{S_d \in S_d} \mathbb{P}_{S_d}(\bigcap_{i<T_d} \{ r_d(X_i^d) < -\epsilon \}) = o(1)
\]
for \( \delta < 1 \). Set
\[
\sigma_d^2 = \|W_1^d\|^2 \frac{\|X_0^d\|^2}{d} = \sigma^2 y.
\]
where \( \sigma = \|W_1^d\| \). We use the same filtration as in (4.11). Let
\[
\begin{cases}
\chi_k^d := r_d(X_k^d) - r_d(X_{k-1}^d), \\
\tilde{X}_k^d := \chi_k^d 1_{\{\|W_k^d\| < \delta^3, r_d(X_{k-1}^d) \in K(\delta)\}}.
\end{cases}
\]
As in Lemma 4.3 it is not difficult to show that the following convergences are sufficient for our claim:
\[
\mathbb{P}_{S_d}(\inf_{0<i<T_d} \chi_k^d 1_{\{\|W_i^d\| \geq \delta^3, r_d(X_{i-1}^d) \in K(\delta)\}} < 0) = o(1), \quad \text{(4.25)}
\]
where the convergence should be uniform for \( S_d \in S_d \). As in Lemma 4.3 the left-hand side of (4.25) is bounded above by
\[
d^{2-\alpha} \mathbb{P}_{S_d}(\chi_1^d < 0, \|W_1^d\| \geq \delta^3, y \in K(\delta)) \leq d^{2-\alpha} \mathbb{P}_{S_d}(Z^d < 0, \|W_1^d\| \geq \delta^3, y \in K(\delta)) \leq d^{2-\alpha} \mathbb{E}_{S_d}\left[ U_1^d \left( \frac{2 - \beta}{d^{1+1/2}} \right)^{\frac{2 - \beta}{\sigma}} \right] = d^{2-\alpha} (2\delta^{-1/2})^{\frac{2 - \beta}{\sigma}} \mathbb{E}[U_1^d] \overset{d^{2-\alpha} (2\delta^{-1/2})^{\frac{2 - \beta}{\sigma}} \mathbb{E}[U_1^d]}{= o(1)}
\]
where we used (2.2) in the last equality. Let \( Z \sim N_\sigma \). Then by Lemma 2.3 and Lemma 4.5 \( 2^{-1} \mathbb{E}_{S_d}[\tilde{B}_d] \) is
\[
d^{1-\alpha} \mathbb{E}_{S_d}[\|Y_{d, \sigma}^d\|_{\mathfrak{h}^d(y)} : \sigma < \delta^3, y \in K(\delta)]
\leq d^{1-\alpha} \mathbb{E}_{S_d}\left[ \left( h_d(Z^d, y) - h(Z^d) \right) + h(Z^d) - h(Z) \right] : \sigma < \delta^3, y \in K(\delta) \]
\leq d^{1-\alpha} \delta^{-1} \left\{ \sup_{y \in K(\delta)} \|h_d(y) - h\|_\infty + \sup_{\sigma < \delta^3} \frac{3\sigma}{d - 1} \right\}
\leq d^{1-\alpha} \delta^{-1} \left\{ \frac{M^d(\delta)}{d} + \frac{3\delta^3}{d - 1} \right\} = o(1).
\]
Finally we check the convergence of $\tilde{C}_d$. Note that if $\sigma = \|W^d\| < d^3$ and $y = \|X^d_0\|^2/d \in K(\delta)$, then

$$
\|X^d_0\|^2 \leq (\|X^d_0\| + \|W^d\|)^2 \leq (\sqrt{\delta d^{-1}} + d^3)^2 \leq (\sqrt{\delta d^{-1}} + \sqrt{d d^{-1}})^2 \leq 4d\delta^{-1}.
$$

(4.26)

For this fact, $d^{\alpha - 2}E_{x,y}[\tilde{C}_d]$ is

$$
sup_{s_d \in \tilde{S}_d} E_{x,y}[\chi^{d^2}, \sigma < d^3, y \in K(\delta)]
= sup_{s_d \in \tilde{S}_d} E_{x,y}[\{2, \|X^d_0\|^2 \leq \delta^{-1}\} + sup_{s_d \in \tilde{S}_d} E_{x,y}[\{2, \|X^d_0\|^2 \leq \delta^{-1}\}]
\leq 2sup_{s_d \in \tilde{S}_d} \{\{2, \|X^d_0\|^2 \leq \delta^{-1}\}
\leq 2\delta^{-1}
$$

where we used reversibility in the forth equation. Then as in the proof of Lemma 4.3 we have

$$
d^{\alpha - 2}E_{x,y}[\tilde{C}_d] \leq E_{x,y}\left[(U^{d^2})^2 \left(\frac{\|X^d_0\|^2}{d}\right), \|X^d_0\|^2 \leq \delta^{-1}\right] \leq 4\delta^{-1}
$$

which proves $\tilde{C}_d = o_{p_{x,y}}(1)$. Hence the claim follows.

$\square$

**Proof of Theorem 3.3.** The proof follows the same lines as the proof of Theorem 3.1 and is omitted.$\square$

### 4.6 Proof of Proposition 3.4 and Theorem 3.4

We need the estimates of

$$
\begin{align*}
N^d(\varepsilon) &= d sup_{y \in K(\varepsilon)} \|\alpha_d(\cdot; y) - \alpha\|_{\infty}, \\
N^d(\varepsilon) &= d sup_{y \in K(\varepsilon)} \|\partial \alpha_d(\cdot; y) - \alpha'\|_{\infty}.
\end{align*}
$$

Under Assumption 1 for any $\varepsilon < 1, lim sup_{d \to \infty} N^d(\varepsilon) < \infty$ and lim sup$_{d \to \infty} N^d(\varepsilon) < \infty$. The proof comes from the same lines as the proof of Lemma 4.5 and is omitted.

**Proof of Proposition 3.4.** The proof is almost the same as that of Proposition 3.2. Set

$$
\begin{align*}
a_d(x, y) &= dE_{x,y}[\pi_k(X^d_1) - \pi_k(X^d_0)], \\
b_d(x, y) &= dE_{x,y}[\pi_k(X^d_1 - \pi_k(X^d_0)^2)], \\
c_d(x, y) &= dE_{x,y}[\pi_k(X^d_1 - \pi_k(X^d_0)^4)].
\end{align*}

(4.27)

Assume $y \in K(\varepsilon) = [\epsilon, 1 - \epsilon]$ throughout ($\epsilon \in (0, 1)$). Then by (3.1)

$$
a_d(x, y) = dy^{1/2}E_{x,y}[\pi_k(W^d)\alpha_d(Z^d; y)]
= y^{1/2}E_{x,y}[\pi_k(X^d_0 + W^d)\partial \alpha_d(Z^d; y)]
= y^{1/2}E_{x,y}[\partial \alpha_d(Z^d; y)] + O(d^{-1/2})
= 1^2E_{x,y}[\partial \alpha_d(Z^d; y)] + O(d^{-1/2})
$$

where we used $\|\partial \alpha_d(\cdot; y)\|_{\infty} \leq \|\alpha'\|_{\infty} + N^d(\varepsilon)/d$ in the third equality. By using the boundedness of $N^d(\varepsilon)$ together with the dominated convergence theorem, we have

$$
E_{x,y}[\partial \alpha_d(Z^d; y)] = E_{x,y}[\alpha'(Z)] + o(1) = -\frac{1}{2} \mu_0(1) + o(1)
$$

20
uniformly for \( y \in K(\varepsilon) \) where \( Z \sim N_1 \). For the diffusion term, by Stein’s identity \( A.2 \),
\[
dE_{x,y}[(\pi_k(W_1^d)^{\otimes 2} - \frac{1}{d} I_k)\alpha_d(Z^d; y)] = \frac{1}{d} E_{x,y} [\pi_k(X_0^d + W_1^d) \otimes \pi_k(W_1^d) \partial_z \alpha_d(Z^d; y)] = o(1)
\]
by the estimate of \( \|\partial_z \alpha_d(\cdot; y)\|_\infty \). By boundedness of \( N^d(\varepsilon) \), we also have
\[
P^x I_k E_{x,y}[\alpha_d(Z^d; y)] = \frac{1}{d} E_{x,y} [\alpha(Z^d)] + o(1) = \frac{1}{d} \mu_0(I) I_k.
\]
Thus we have
\[
b_d(x, y) = dy E_{x,y} [\pi_k(W_1^d)^{\otimes 2}] = y P^x \mu_0(I) I_k + o(1).
\]
Finally,
\[
c_d(x, y) = dE_{x,y} [\|\pi_k(W_1^d)\|^4] = o(1).
\]

Therefore by the convergences of \( B.19 \) and \( 4.27 \), the step Markov process \( (r_d(X_{[d]}^d)), \pi_k(X_{[d]}^d))_{t \geq 0} \) converges to \( (y, Y_2^d))_{t \geq 0} \) where \( Y_2^d \) is defined in Proposition \( 3.3 \) and \( y \sim Q \) by Theorem IX 4.21 of \textit{Jacod and Shiryaev} 2003. Thus the claim follows.

\textbf{Proof of Theorem 3.4.} Weak consistency of RWM\((P_d, S_d)\) is a direct conclusion from Lemma \( 3.2 \) and Proposition \( 3.3 \). We prove the optimality.

Suppose now that RWM\((P_d, S_d)\) is weakly consistent with the rate \( T_d = d^{2 - \alpha} \). Let \( g \) be any bounded continuous function on \( \mathbb{R} \). Then applying this to \( 2.5 \), we have
\[
E \left[ \left| M_d^{-1} \sum_{m=0}^{M_d-1} g(\pi_1(X_m^d)) - P_1(g) \right| \right] = o(1).
\]

Thus by using the filtration \( 4.11 \), we have
\[
o(1) = \left| \left| M_d^{-1} \sum_{m=0}^{M_d-1} g(\pi_1(X_m^d)) - P_1(g) \right| \right| = \left| \left| M_d^{-1} \sum_{m=0}^{M_d-1} \left| g(\pi_1(X_m^d)) \right| F_m^d - P_1(g) \right| \right|
\]
where \( \bigvee_{m \in \mathbb{N}_0} F_m^d \) is the smallest \( \sigma \)-algebra that includes \( \bigcup_{m \in \mathbb{N}_0} F_m^d \). Since the target distribution \( P_d \) is a rotationally symmetric distribution, \( E[g(\pi_1(X_m^d)))|F_m^d] = E_{y=r_d(X_m^d)}[g(y^{1/2}U_1^d)] \) where \( E_y[\cdot] \) is the conditional expectation defined in \( 4.2 \) and \( U_1^d \) is the first component of the uniform distribution on the surface \( \{ x \in \mathbb{R}^d; \|x\| = d \} \). Therefore
\[
\left| \left| M_d^{-1} \sum_{m=0}^{M_d-1} \left| E_{y=r_d(X_m^d)}[g(y^{1/2}U_1^d)] - P_1(g) \right| \right| \right| = o_T(1).
\]

By Lemma \( 2.8 \) we can replace \( E_{y}[g(y^{1/2}U_1^d)] \) by \( E[g(y^{1/2}W)] \) where \( W \sim N(0, 1) \). Therefore for any bounded continuous function \( g \), the convergence \( 2.6 \) holds for \( f(y) = E[g(y^{1/2}W)] \). Set \( M_d = d^{2 - \alpha/2} \).

On the other hand, by Lemma \( 4.6 \), \( M_d^{-1} \sum_{m=0}^{M_d-1} f(r_d(X_m^d)) = f(r_d(X_0^d)) + o_T(1) \). Thus \( 2.6 \) is satisfies only if \( f(r_d(X_0^d)) - P_1(g) = f(Y_0^d) - P_1(g) = o_T(1) \) where \( Y_0^d \sim Q \). Since \( Q \) has a derivative \( q(y) > 0 \), it is possible only if \( f \) is a constant. This contradicts weak consistency of RWM\((P_d, S_d)\) with the rate \( T_d = d^{2 - \alpha} \) and hence the claim follows.

\textbf{Acknowledgements.}

I am grateful to Hiroki Masuda and Masaaki Fukasawa for many helpful insights and suggestions.
A \hspace{1em} Stein’s method and Stein’s characteristics

For calculation of normal distribution, we will use few techniques from Stein’s method. This method provides a simple way to evaluate complicated calculation related to the normal distribution. See \cite{Chen et al. 2011} for the details. Write \( N = N(0, 1), \) \( \phi(z) = \exp(-z^2/2)/\sqrt{2\pi} \) and \( \Phi(z) = \int_{-\infty}^{z} \phi(w)dw. \) We remark Stein’s characterization of the normal distribution; \( Z \sim N \) if and only if

\[
E[f'(Z)] = E[Zf(Z)] \tag{A.1}
\]

for any absolutely continuous function \( f(z). \) Using \( zf(z) \) in place of \( f \) in the above, we obtain

\[
E[Zf'(Z)] = E[(Z^2 - 1)f(Z)]. \tag{A.2}
\]

Moreover, for any \( N \)-integrable function \( h(z), \) there is a solution of the ordinary differential equation (called Stein’s equation)

\[
f'(z) - zf(z) = h(z) - Nh \tag{A.3}
\]

which is given by \( f(z) = \mathcal{G}(h, N) := \phi(z)^{-1} \int_{-\infty}^{z} (h(w) - Nh)\phi(w)dw. \) It satisfies

\[
\begin{align*}
||f||_{\infty} & \leq \min \{ \sqrt{2\pi}||h||_{\infty}, 2||h'||_{\infty} \}, \\
||f'||_{\infty} & \leq \min \{ 4||h||_{\infty}, \sqrt{2/\pi}||h'||_{\infty} \}, \\
||f''||_{\infty} & \leq 2||h'||_{\infty}.
\end{align*}
\]

In this paper, we will focus on \( N_{\sigma} = N(\sigma^2/2, \sigma^2). \) For this measure, \( Z \sim N_{\sigma} \) if and only if

\[
\sigma^2 E[f'(Z)] = E[(Z - \sigma^2/2)f(Z)] \tag{A.4}
\]

and the solution of Stein’s equation

\[
\sigma^2 f'(z) - (z - \sigma^2/2)f(z) = h(z) - N_{\sigma}h \tag{A.5}
\]

is given by

\[
f(z) = \mathcal{G}(h, N_{\sigma})(z) := \sigma^2\phi_{\sigma}(z)^{-1} \int_{-\infty}^{z} (h(w) - N_{\sigma}h)\phi_{\sigma}(w)dw
\]

where \( \phi_{\sigma} \) is the probability distribution function of \( N_{\sigma}. \) This solution can be written as \( f(z) = \sigma^{-1}f_0((z - \sigma^2/2)/\sigma) \) where \( f_0 = \mathcal{G}(h(\sigma^2/2 + \sigma.), N) \) is the solution of \( \text{(A.3)} \) for \( h(2z/\sigma) + \sigma z). \) By this fact,

\[
\begin{align*}
||f||_{\infty} & \leq \min \{ \sqrt{2\pi\sigma^{-1}}||h||_{\infty}, 2||h'||_{\infty} \}, \\
||f'||_{\infty} & \leq \min \{ 4\sigma^{-2}||h||_{\infty} \sqrt{2/\pi}\sigma^{-1}||h'||_{\infty} \}, \\
||f''||_{\infty} & \leq 2\sigma^{-2}||h'||_{\infty}.
\end{align*}
\]

An important property for \( N_{\sigma} \) is the following Girsanov’s formula: if \( Z \sim N_{\sigma}, \) then for any bounded measurable function \( f, \)

\[
E[f(Z), Z < 0] = E[f(-Z)e^{-Z}, Z > 0]. \tag{A.7}
\]

This formula is used throughout in this paper. For example, by using this, \( N_{\sigma}h = E[h(Z)] = 0 \) for \( h(z) = ze^{-z^2}. \) Using this formula together with Stein’s method, we obtain the explicit form of positive sequences \( \mu_k(\sigma) = E[Z^ke^{-Z}] \) \((k \in \mathbb{N}_0)\) where \( Z \sim N_{\sigma}. \)

**Lemma A.1.** Let \( \sigma > 0. \) We have \( \mu_k(\sigma) = 2(-1)^kE[Z^k, Z < 0] \) and

\[
\begin{align*}
\mu_0(\sigma) &= 2\Phi(-\sigma/2), \\
\mu_1(\sigma) &= -\sigma^2\Phi(-\sigma/2) + 2\sigma\phi(-\sigma/2),
\end{align*}
\]
and
\[ \mu_k(\sigma) = -\frac{\sigma^2}{2} \mu_{k-1}(\sigma) + (k-1)\sigma^2 \mu_{k-2}(\sigma) \quad (k \geq 2). \]  
(A.8)

In particular, \( \sup_{\sigma > 0} \sigma^l \mu_k(\sigma) < \infty \) for any \( k, l \in \mathbb{N}_0 \). Moreover, for \( k \geq 1 \),
\[ \mu'_k(\sigma) = \frac{k}{\sigma} \mu_k(\sigma) - \frac{k\sigma}{2} \mu_{k-1}(\sigma). \]

Proof. By Girsanov’s formula (A.7), \( \mathbb{E}[|Z^k| \exp(-Z^+)] = 2(1)^k \mathbb{E}[Z^k, Z < 0] \) and hence
\[ \mu_0(\sigma) = \mathbb{E}[\exp(-Z^+)] = 2\mathbb{P}(Z < 0) = \Phi(-\frac{\sigma}{\mu}), \]

and
\[ \mu_1(\sigma) = -2\mathbb{E}[Z, Z < 0] = -2 \int_{-\infty}^{-\sigma^2/2} (\frac{\sigma^2}{2} + \sigma w) \phi(w) dw = -\sigma^2 \Phi(-\frac{\sigma}{\mu}) + 2\sigma \phi(-\frac{\sigma}{\mu}). \]

For \( k \geq 2 \), by definition,
\[ \mu_k(\sigma) = 2(1)^k \mathbb{E}[Z^k, Z < 0] = 2(1)^k \mathbb{E}\left[ (Z - \frac{\sigma^2}{2}) Z^{k-1}, Z < 0 \right] \]
\[ = 2(1)^k \mathbb{E}\left[ (Z - \frac{\sigma^2}{2}) Z^{k-1}, Z < 0 \right] + \sigma^2(1)^k \mathbb{E}[Z^{k-1}, Z < 0] \]
\[ = (k-1)\sigma^2 2(1)^k \mathbb{E}[Z^{k-2}, Z < 0] + \sigma^2(1)^k \mathbb{E}[Z^{k-1}, Z < 0] \]
\[ = (k-1)\sigma^2 \mu_{k-2}(\sigma) - \frac{\sigma^2}{2} \mu_{k-1}(\sigma), \]

where we used (A.4) in the third equality. The uniformly bounded property follows from \( \lim_{\sigma \to \infty} \sigma^l \phi(-\sigma/2) = 0 \) and \( \lim_{\sigma \to \infty} \sigma^l \Phi(-\sigma/2) = 0 \) for any \( l \in \mathbb{N} \). For the last claim, we need the following representation:
\[ \mu_k(\sigma) = 2(1)^k \mathbb{E}[Z^k, Z < 0] = (1)^k 2 \int_{-\infty}^{-\frac{\sigma^2}{2}} (\frac{\sigma^2}{2} + \sigma w)^k \phi(w) dw. \]

Then by differentiating both sides, we obtain
\[ \mu'_k(\sigma) = -(1)^k 2k \int_{-\infty}^{-\frac{\sigma^2}{2}} (\frac{\sigma^2}{2} + \sigma w)^k \phi(w) dw \]
\[ = -(1)^k 2k \int_{-\infty}^{-\frac{\sigma^2}{2}} \left\{ \sigma^2 \phi(w)^k + \frac{\sigma^2}{2} \phi(w)^k \right\} dw \]
\[ = \frac{k}{\sigma} \mu_k(\sigma) - \frac{k\sigma}{2} \mu_{k-1}(\sigma). \]

\[ \Box \]

Lemma A.2. Let \( f = \mathfrak{h}(h, N_\sigma) \) for \( h(z) = z e^{-z^+} \). Then \( N_\sigma(f) = -\sigma^{-2} \mu_2(\sigma), N_\sigma(f') = \sigma^{-2} \mu_2(\sigma)/2 \) and \( N_\sigma(f'') = \sigma^{-2}(\mu_1(\sigma) - \mu_2(\sigma))/3 \).

Proof. Write \( \mu_k \) for \( \mu_k(\sigma) \). By \( N_\sigma h = 0 \), the solution of Stein’s equation for \( h(z) = z e^{-z^+} \) is
\[ f(w) = \sigma^{-2} \phi_\sigma(w)^{-1} \mathbb{E}[Ze^{-Z^+}, Z \leq w] = \sigma^{-2} \phi_\sigma(w)^{-1} \mathbb{E}[Z, Z \leq -|w|] \]
for \( Z \sim N_\sigma \) where the second equation follows from (A.7). Then by definition,
\[ \mathbb{E}[Z^k f(Z)] = \sigma^{-2} \int w^k \mathbb{E}[Z, Z \leq -|w|] dw. \]
Further more, we assume the following. Corresponding to Lemma B.1. Let and each $X \rightarrow \infty$. Since $w \rightarrow E[Z, Z \leq -|w|]$ is a symmetric function, we have $E[Z^k f(Z)] = 0$ if $k$ is odd, and if $k$ is even,

$$E[Z^k f(Z)] = 2\sigma^{-2} \int_0^\infty w^k E[Z, Z \leq -|w|] dw = 2\sigma^{-2} E[-\frac{Z^{k+2}}{k+1}, Z \leq 0] = -\sigma^{-2} \frac{\mu_k+2}{k+1}.$$ 

Therefore for $k = 0$, we have $N_\sigma f = -\sigma^{-2} \mu_2$. By this fact, together with (A.4), we have

$$N_\sigma(f') = \sigma^{-2} E[(Z - \frac{\sigma^2}{2}) f(Z)] = \sigma^{-2} (0 - \frac{\sigma^2}{2} N_\sigma f) = \sigma^{-2} \mu_2/2.$$ 

The integral of $f''$ is, by applying $(z - \sigma^2/2) f(z)$ to (A.3) in place of $f$,

$$N_\sigma(f'') = \sigma^{-2} E[(Z - \frac{\sigma^2}{2}) f'(Z)] = \sigma^{-4} E[\{(Z - \frac{\sigma^2}{2})^2 - \sigma^2\} f(Z)]$$

$$= \sigma^{-4} E[(Z^2 - \sigma^2 Z + (\frac{\sigma^4}{4} - \sigma^2)) f(Z)]$$

$$= \sigma^{-4} \{(-\sigma^{-2} \mu_4) - 0 + (\frac{\sigma^4}{4} - \sigma^2) (-\sigma^{-2} \mu_2)\}.$$ 

By (A.8), simple calculation yields

$$\mu_4 = -\sigma^4 \mu_1 + (\frac{\sigma^4}{4} + 3\sigma^2) \mu_2.$$ 

Then

$$N_\sigma(f'') = \frac{\sigma^{-2}}{3} (\mu_1 - \mu_2). \quad \square$$

## B Some properties of consistency

Suppose that $(X_t)_{t \in [0, \infty)}$ is a (strictly) stationary process and set $f : \mathbb{R} \rightarrow [-1, 1]$. Then for $S \leq T$, we have

$$\left| E\left[\frac{1}{T} \int_0^T f(X_t) dt\right]\right| \leq \left| E\left[\frac{1}{S} \int_0^S f(X_t) dt\right]\right| + \frac{S}{T}. \quad (B.1)$$ 

This is an easy consequence from $T^{-1} \int_0^T f(X_t) dt = T^{-1} \sum_{i=0}^{T/S-1} \int_{iS}^{iS+S} f(X_t) dt = T^{-1} \int_{[T/S]}^T f(X_t) dt$ with $x[x^{-}] \leq 1$. Using this, we can prove the following. See Lemma 2.4 of [Kamatani 2013] for the proof.

**Lemma B.1.** Let $\{X^d_m; m = 0, 1, \ldots\}$ be a sequence of stationary Markov chain with invariant probability measure $P_d$. Set $Y^d_t = X^d_{[Ta]}$ for some $T_d \rightarrow \infty$. If $Y^d$ converge in law to a stationary ergodic process, then

$$\frac{1}{M_d} \sum_{m=0}^{M_d-1} f(X^d_m) - P_d(f) = o_p(1)$$

for any $M_d \rightarrow \infty$ such that $M_d/T_d \rightarrow \infty$ for any bounded and continuous function $f$.

We need a slightly generalization of this lemma. For our purpose, $X^d_m$ has two parts, $X^d_m = (X^d_m, X^d_m)$ and each $X^d,1 = (X^d_m)_{m \in \mathbb{N}_0}$ and $X^d,2 = (X^d_m)_{m \in \mathbb{N}_0}$ has different convergence rate $T_d^1$ and $T_d^2$ with respective. Corresponding to $X^d,1$ and $X^d,2$, the invariant probability measure has the following decomposition

$$P_d(dx_1 dx_2) = P^1_d(dx_1) P^2_d(dx_2 | x_1).$$ 

Further more, we assume the following.
Assumption 2. 1. Positive sequences $T^1_d, T^2_d \to \infty$ but $T^2_d/T^1_d \to 0$.

2. For $Y^d_{t} = X^d_{[t/T_d]}$, $Y^d_{1} \to Y^1_{1}$ where $Y^1_{1}$ is stationary and ergodic continuous process with the invariant probability measure $P^1$.

3. For $Y^d_{t} = X^d_{[t/T_d]}$, $Y^d_{1}$ converges to $Y = (Y_t)_{t \in [0, \infty)} = (X^1_{0}, Y^2_{0}(X^1_{0}))$ where $Y^2_{1}(x) = (Y^2_{t}(x))_{t \in [0, \infty)}$ is a stationary and ergodic process with the invariant probability measure $P^{2\mid 1}(\cdot|x)$ for each $x$, and $X^1_{0} \sim P^1$.

4. For any bounded continuous function $f$, $P^{2\mid 1} f(X_{1}) = \int f(x_1, x_2) P^{2\mid 1}(dx_2|x_1)$ is continuous in $x_1$.

Lemma B.2. Under the above assumption,

$$
\frac{1}{M_d} \sum_{m=0}^{M_d-1} f(X^d_m) - P_d(f) = o_P(1)
$$

for any continuous and bounded function $f$ and for $M_d \to \infty$ such that $M_d/T^1_d \to \infty$.

Proof. First we remark that $P_d \to P$ where $P(dx_1 dx_2) = P^1(dx_1) P^{2\mid 1}(dx_2|x_1)$. By triangular inequality,

$$
\begin{align*}
\mathbb{E} \left[ \left| \frac{1}{M_d} \sum_{m=0}^{M_d-1} f(X^d_m) - P_d(f) \right| \right] & \leq \mathbb{E} \left[ \left| \frac{1}{M_d} \sum_{m=0}^{M_d-1} f(X^d_m) - (P^{2\mid 1} f)(X^d_{1}) \right| \right] \\
& \quad + \mathbb{E} \left[ \left| \frac{1}{M_d} \sum_{m=0}^{M_d-1} (P^{2\mid 1} f)(X^d_{m}) - P_d(f) \right| \right] =: A_d + B_d.
\end{align*}
$$

Then $B_d = o(1)$ by applying Lemma [B.1] for $g := P^{2\mid 1} f$ in place of $f$.

Next we prove the convergence of $A_d$. By (B.1), without loss of generality, we can assume

$$
\frac{M_d}{T^1_d} \to \infty, \quad \text{and} \quad \frac{M_d}{T^2_d} \to 0.
$$

By (B.1) again, for any $S$ and for sufficiently large $d$ so that $ST^2_d \leq M_d$, we have

$$
A_d \leq \mathbb{E} \left[ \left| \frac{1}{ST^2_d} \sum_{m=0}^{ST^2_d-1} f(X^d_m) - g(X^d_{m+1}) \right| ight] + o(1),
$$

$$
= \mathbb{E} \left[ \left| \frac{1}{S} \int_{[0, S]} f(Y^d_t)dt - g(X^d_{0}) \right| \right] + o(1),
$$

$$
= \mathbb{E} \left[ \left| \frac{1}{S} \int_{[0, S]} f(Y^d_t)dt - g(X^d_{0}) \right| \right] + o(1).
$$

Note here that continuity of $Y^1_{1}$ is used to show $\sup_{m \leq ST^2_d} |g(X^d_{m+1}) - g(X^d_{m})| = o_P(1)$. Then by taking $S \to \infty$, the right hand side converges to 0 and hence the claim follows.

C Proof of technical lemmas

C.1 Proof of Lemma 2.2

Proof of Lemma 2.2 Set $\epsilon > 0$ and set $K(\delta) = [\delta, \delta^{-1}]$ for $\delta \in (0, 1)$. The probability distribution function $q_d(y)$ is

$$
q_d(y) = \int_0^\infty g(z; d/2, d/2)q(z)\frac{dz}{y},
$$

25
Set \( g_d(z) := g(z; d/2, d/2) \). We divide \( q_d(y) - q(y) \) into \( a_d(y) + b_d(y) \) where

\[
a_d(y) = \int_0^\infty g_d(z)1_{K(\delta)}(z)\left(\frac{q(z)}{z} - q(y)\right)dz, \quad b_d(y) = \int_0^\infty g_d(z)1_{K(\delta)}(z)\left(\frac{q(z)}{z} - q(y)\right)dz.
\]

The family of functions \( \{g_d(z)\}_{d \in \mathbb{N}} \) works as a mollifier. Since \( \Gamma(d + \alpha)/\Gamma(d)d^\alpha \to 1 \), we have

\[
\frac{g_{d+1}(z)}{g_d(z)} = \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d}{2} + 1\right)}\left(\frac{d}{2} + 1\right)^{1/2}\left(1 + \frac{1}{d}\right)^{d/2}z^{-1/2}e^{-z/2} \to z^{-1/2}e^{-(z-1)/2}
\]

uniformly in \( z \). The right-hand side is less than 1 if \( z \neq 1 \), and hence for any \( z \in K(\delta)^c \), \( g_d(z) \) decreases as \( d \) increases when \( d \) is sufficiently large. Thus by the dominated convergence theorem, we have

\[
\|a_d\|_\infty \leq \|q\|_\infty \int_0^\infty g_d(z)1_{K(\delta)^c}(z)\frac{dz}{\min\{z, 1\}} \to 0.
\]

Set \( \tau \in (0, 1/2) \) and \( \delta \in (1/2, 1) \) so that \( K(\delta) \subset K(1/2) \). Then

\[
\|b_d1_{K(\tau)^c}\|_\infty \leq \sup_{y \in K(\tau)^c, z \in K(\delta)} \left| \frac{q(y)}{z} - q(y) \right| \leq \sup_{y \in K(\tau)^c, z \in K(1/2)} \left| \frac{q(y)}{z} - q(y) \right| \leq 2\|q1_{K(2\tau)^c}\|_\infty.
\]

Choose \( \tau \in (0, 1/2) \) so that the right-hand side is bounded above by \( \epsilon/3 \). On the other hand,

\[
\|b_d1_{K(\tau)}\|_\infty \leq \sup_{y \in K(\tau), z \in K(\delta)} \left| \frac{q(y)}{z} - q(y) \right| \\
\leq \sup_{y \in K(\tau), z \in K(\delta)} \left| \frac{q(y)}{z} - q(y) \right| + \sup_{y \in K(\tau), z \in K(\delta)} \left| \frac{q(y)}{z} - q(y) \right| \\
\leq 2\sup_{|x-y| \leq \tau^{-1}(\delta^{-1} - 1)} \|q(x) - q(y)\| + (\delta^{-1} - 1)\|q\|_\infty.
\]

By uniform continuity of \( q \), we can choose \( \delta \in (1/2, 1) \) so that the both two terms in the right hand side is smaller than \( \epsilon/3 \). Thus \( \|b_d\|_\infty \leq \epsilon \). By taking \( \epsilon \to 0 \), \( \|q_d - q\|_\infty \leq \|a_d\|_\infty + \|b_d\|_\infty \to 0 \). The proof of \( \|q_d^{(i)} - q_i^{(i)}\|_\infty \to 0 \) (\( i = 1, 2 \)) follows the same line and is omitted. \( \square \)

### C.2 Proof of Lemma 4.5

The proof of the following is quite easy.

**Lemma C.1.** For \( z > 0, \beta > \alpha > 0 \),

\[
z^\alpha (1 + z)^{-\beta} \leq \left(\frac{\alpha}{\beta - \alpha}\right)^\alpha, \quad z^\alpha e^{-z} \leq \alpha^\alpha e^{-\alpha}.
\]

**Proof of Lemma 4.5** Set

\[
\begin{align*}
\alpha(z) &= e^{-z^+}, \\
\tilde{\alpha}_d(z; y) &= \frac{q_d(y(1 + \frac{2z}{d}))}{q_d(y)}e^{-z^+}, \\
\alpha_d(z; y) &= \frac{q_d(y(1 + \frac{2z}{d}))}{q_d(y)}(1 + \frac{2z}{d})\frac{2z}{d}.
\end{align*}
\]

Since \( \alpha_d(z; y) = \tilde{\alpha}_d(z; y) = \alpha(z) = 1 \) for \( z < 0 \), we assume \( z > 0 \) throughout in this proof. We also assume \( y \in K(\epsilon) \). By Taylor’s expansion, we have

\[
\begin{align*}
0 &\leq 1 - (1 + \frac{2z}{d})^{d/2}e^{-z} = 1 - \exp\left(-z + \frac{d}{2} \log(1 + \frac{2z}{d})\right) \\
&\leq z - \frac{d}{2} \log(1 + \frac{2z}{d}) \leq \frac{z^2}{d}.
\end{align*}
\]
where we used \(1 - e^z \leq -z\) in the second inequality, and \(\log(1 + z) \geq z - z^2/2\) in the last inequality. By this,

\[
|(1 + \frac{2z}{d})^{2z/d} - e^{-z}| \leq |(1 + \frac{2z}{d})^{2z/d} - (1 + \frac{2z}{d})^{-\frac{d}{2}}| + |(1 + \frac{2z}{d})^{-\frac{d}{2}} - e^{-z}|
\]

\[
\leq \frac{2z + z^2}{d}(1 + \frac{2z}{d})^{-\frac{d}{2}}.
\]

Using this, we have

\[
|\alpha_d(z; y) - \tilde{\alpha}_d(z; y)| \leq \frac{2z + z^2}{d}(1 + \frac{2z}{d})^{-\frac{d}{2}} \frac{\|q_d\|_{\infty}}{\inf_{K(e)} q_d},
\]

\[
|\tilde{\alpha}_d(z; y) - \alpha(z)| \leq \frac{2z^2 e^{-1}}{d} \frac{\|q'_d\|_{\infty}}{\inf_{K(e)} q_d}.
\]

By these estimates together with Lemma \[C.1\] we obtain

\[
|h_d(z; y) - h(z)| \leq z|\alpha_d(z; y) - \tilde{\alpha}_d(z; y)| + z|\tilde{\alpha}_d(z; y) - \alpha(z)|
\]

\[
\leq \frac{2z^2 + z^3}{d}(1 + \frac{2z}{d})^{-\frac{d}{2}} \frac{\|q_d\|_{\infty}}{\inf_{K(e)} q_d} + \frac{2z^2}{d} \frac{e^{-1}\|q'_d\|_{\infty}}{\inf_{K(e)} q_d}
\]

\[
\leq \left(\frac{d}{2} \left(\frac{2}{d/2 - 2}\right)^2 + \frac{d^2}{2^3} \left(\frac{3}{d/2 - 3}\right)^3\right) \frac{\|q_d\|_{\infty}}{\inf_{K(e)} q_d} + \frac{8e^{-2} e^{-1}\|q'_d\|_{\infty}}{d^{\frac{d}{2}} \inf_{K(e)} q_d}
\]

\[
\leq \frac{d^{-1}}{d} \left(\frac{32 + 63}{d} \frac{\|q_d\|_{\infty}}{\inf_{K(e)} q_d} + \frac{8e^{-2} e^{-1}\|q'_d\|_{\infty}}{d^{\frac{d}{2}} \inf_{K(e)} q_d}\right)
\]

for \(d \geq 12\) where we used \(d/2 - 2 > d/2 - 3 > d/4\). Thus \(\limsup M^d(\epsilon) < \infty\) by Lemma \[2.2\] On the other hand, since \((1 + 2z/d)^{(2-d)/2}' = \frac{1-2/d}{1+2z/d} (1 + 2z/d)^{(2-d)/2}\), we have

\[
\partial_z h_d(z; y) - h'(z) = (\alpha_d(z; y) - \alpha(z)) + z(\partial_z \alpha_d(z; y) - \alpha'(z))
\]

\[
= (\alpha_d(z; y) - \alpha(z)) - \left\{ \frac{1-2/d}{1+2z/d} \partial_z h_d(z; y) - h(z) \right\}
\]

\[
+ \frac{2z y}{d} \frac{(1 + \frac{2z}{d})^{2-d/2}}{q_d(y)} q'_d(y(1 + 2z/d)),
\]

\[
= A_d(z; y) - B_d(z; y) + C_d(z; y).
\]

As in the estimate of \(\|h_d - h\|_{\infty}\), we have

\[
|A_d(z; y)| \leq \|\alpha_d(\cdot; y) - \alpha\|_{\infty} \leq d^{-1}(4 + 4^2) \frac{\|q_d\|_{\infty}}{\inf_{K(e)} q_d} + 2e^{-1} \frac{\|q'_d\|_{\infty}}{\inf_{K(e)} q_d}
\]

for \(d/2 - 2 \geq d/2 - 1 > d/4\) and also

\[
|B_d(z; y)| \leq \frac{1-2/d}{1+2z/d} \|h_d(\cdot; y) - h\|_{\infty} + \left(1 - \frac{1-2/d}{1+2z/d}\right) h(z)
\]

\[
\leq \|h_d(\cdot; y) - h\|_{\infty} + d^{-1} (2z + 2) h(z)
\]

\[
\leq d^{-1} (M_d(\epsilon) + 8e^{-2} + 2e^{-1}).
\]

Finally,

\[
|C_d(z; y)| \leq (d/2 - 3)^{-1} \frac{\epsilon^{-1}\|q_d\|_{\infty}}{\inf_{K(e)} q_d} \leq d^{-1} (4 + 2e^{-1}) \frac{\epsilon^{-1}\|q'_d\|_{\infty}}{\inf_{K(e)} q_d}
\]

if \(d \geq 12\) since \(d/2 - 3 \geq d/4\). Thus

\[
M^d(\epsilon) \leq M^d(\epsilon) + 8e^{-2} + 2e^{-1} + 2d \frac{\epsilon^{-1}\|q_d\|_{\infty}}{\inf_{K(e)} q_d} + (4 + 2e^{-1}) \frac{\epsilon^{-1}\|q'_d\|_{\infty}}{\inf_{K(e)} q_d}
\]

Hence the claim follows.
References

Bo Martin Bibby, Ib Michael Skovgaard, and Michael Sørensen. Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli*, 11(2):191–220, 2005. ISSN 1350-7265. doi: 10.3150/bj/1116340291.

Sourav Chatterjee and Elizabeth Meckes. Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 4:257–283, 2008. ISSN 1980-0436.

Louis H.Y. Chen, Larry Goldstein, and Qi-Man Shao. *Normal Approximation by Stein’s Method*. Probability and Its Applications. Berlin, Heidelberg : Springer Berlin Heidelberg, 2011., 2011. ISBN 9783642150074.

Persi Diaconis and David Freedman. A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist.*, 23(2, suppl.):397–423, 1987. ISSN 0246-0203.

A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient Metropolis jumping rules. In *Bayesian statistics, 5 (Alicante, 1994)*, Oxford Sci. Publ., pages 599–607. Oxford Univ. Press, New York, 1996.

Valentine Genon-Catalot and Jean Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 29(1):119–151, 1993. ISSN 0246-0203.

Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*. Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, Berlin, 2nd edition, 2003.

Søren F. Jarner and Gareth O. Roberts. Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Statist.*, 34(4):781–815, 2007. ISSN 0303-6898. doi: 10.1111/j.1467-9469.2007.00557.x.

Søren F. Jarner and Richard L. Tweedie. Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli*, 9(4):559–578, 2003. ISSN 1350-7265. doi: 10.3150/bj/1066223269.

Søren Fiig Jarner and Ernst Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.*, 85(2):341–361, 2000. ISSN 0304-4149. doi: 10.1016/S0304-4149(99)00082-4.

Kengo Kamatani. Metropolis-Hastings algorithms with acceptance ratios of nearly 1. *Ann. Inst. Statist. Math.*, 61(4):949–967, 2009. ISSN 0020-3157. doi: 10.1007/s10463-008-0180-6.

Kengo Kamatani. Local weak consistency of Markov chain Monte Carlo methods with application to mixture model. *Bulletin of Informatics and Cybernetics*, 45:103–123, 2013. URL [http://arxiv.org/abs/1103.5679](http://arxiv.org/abs/1103.5679).

Kengo Kamatani. Local consistency of Markov chain Monte Carlo methods. *Ann. Inst. Statist. Math.*, 66(1):63–74, 2014. ISSN 0020-3157. doi: 10.1007/s10463-013-0403-3.

Samuel Kotz and Saralees Nadarajah. *Multivariate t distributions and their applications*. Cambridge University Press, Cambridge, 2004. ISBN 0-521-82654-3. doi: 10.1017/CBO9780511550683.

Kerrie L. Mengersen and Richard L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996. ISSN 0090-5364. doi: 10.1214/aos/1033066201.

Peter Neal and Gareth Roberts. Optimal scaling of random walk Metropolis algorithms with non-Gaussian proposals. *Methodol. Comput. Appl. Probab.*, 13(3):583–601, 2011. ISSN 1387-5841. doi: 10.1007/s11009-010-9176-9.

Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling of discrete approximations to langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(1):255–268, 1998. ISSN 1467-9868. doi: 10.1111/1467-9868.00123.
Gareth O. Roberts and Richard L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996. ISSN 0006-3444. doi: 10.1093/biomet/83.1.95.

Gareth O. Roberts, Andrew Gelman, and Walter R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997. ISSN 1050-5164. doi: 10.1214/aoap/1034625254.

Ken-iti Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. ISBN 0-521-55302-4. Translated from the 1990 Japanese original, Revised by the author.

Chris Sherlock and Gareth Roberts. Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets. *Bernoulli*, 15(3):774–798, 2009. ISSN 1350-7265. doi: 10.3150/08-BEJ176.

Luke Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1762, 1994. ISSN 0090-5364. doi: 10.1214/aos/1176325750. With discussion and a rejoinder by the author.