

Efficient strategy for the Markov chain Monte Carlo in high-dimension with heavy-tailed target probability distribution

KENGO KAMATANI

Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan. E-mail: kamatani@sigmath.es.osaka-u.ac.jp

The purpose of this paper is to introduce a new Markov chain Monte Carlo method and to express its effectiveness by simulation and high-dimensional asymptotic theory. The key fact is that our algorithm has a reversible proposal kernel, which is designed to have a heavy-tailed invariant probability distribution. A high-dimensional asymptotic theory is studied for a class of heavy-tailed target probability distributions. When the number of dimensions of the state space passes to infinity, we will show that our algorithm has a much higher convergence rate than the pre-conditioned Crank–Nicolson (pCN) algorithm and the random-walk Metropolis algorithm.

Keywords: Consistency; Malliavin calculus; Markov chain; Monte Carlo; Stein’s method

1. Introduction

The Markov chain Monte Carlo method (MCMC) is a commonly used technique for evaluating complicated integrals, particularly in the high-dimensional state space. Many new methods have been developed over the last few decades. However, it is still very difficult to choose an MCMC that works well for a given function and a given measure, which is called the target probability distribution. The choice of MCMC strongly depends on the tail behavior of the target probability distribution. In particular, it is well known that many MCMC algorithms behave poorly for heavy-tailed target probability distributions.

In our previous work, in [10], we studied some asymptotic properties of the random-walk Metropolis (RWM) algorithm for a class of heavy-tailed target probability distributions. The algorithm has a very slow convergence for this class. Finding a more effective strategy is an important unresolved problem.

A candidate of this algorithm, the pre-conditioned Crank–Nicolson algorithm (pCN), appeared for the first time in [17]. The method is a simple modulation of a classical Gaussian RWM algorithm, and therefore their computational costs are almost identical. The effectiveness of this simple candidate was provided in the simulation by [3] and its theoretical effect was provided in [1,4,21] and [8]. However, our simulation shows that it works only for a specific light-tailed distribution and does not work well otherwise, especially for the class of heavy-tailed target probability distributions considered in this paper. In Theorem 3.1, we will prove its regarding convergence rate.

In this paper, we introduce a new algorithm which is a slight modification of the original pCN algorithm although their performance is completely different. It works well for a large class of target probability distributions. Let us describe our new algorithm, the mixed pre-conditioned Crank–Nicolson algorithm (MpCN). Let $P(dx) = p(x) dx$ be the target probability distribution on \mathbb{R}^d . Fix $\rho \in (0, 1)$. Set initial value $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ and let $\|x\| = (\sum_{i=1}^d x_i^2)^{1/2}$. The algorithm goes as follows:

- Generate $r \sim \text{Gamma}(d/2, \|x\|^2/2)$.
- Generate $x^* = \rho^{1/2}x + (1 - \rho)^{1/2}r^{-1/2}w$ where w follows the standard normal distribution.
- Accept x^* as x with probability $\alpha(x, x^*)$, and otherwise, discard x^* , where

$$\alpha(x, y) = \min \left\{ 1, \frac{p(y)\|x\|^{-d}}{p(x)\|y\|^{-d}} \right\}.$$

Here, $\text{Gamma}(v, \alpha)$ is the Gamma distribution with probability density function $g(x; v, \alpha) = x^{v-1}\alpha^v \exp(-\alpha x) / \Gamma(v)$.

The key fact is that the proposal kernel of the proposed algorithm has the heavy-tailed invariant probability distribution. Thus, it is reasonable that the new method works better than the pCN algorithm for the class of heavy-tailed target probability distributions. In addition, we show by simulation that the new method is at least as good as that of the pCN algorithm, even for the light-tailed target probability distribution. Our method works well for a wide class of target probability distributions.

We study its theoretical properties via high-dimensional asymptotic theory. The high-dimensional asymptotic theory for MCMC was first appeared in [24] and further developed in [25]. See [3] for recent results. We use this framework together with the study of consistency of MCMC by [11].

The main technical tools are the Malliavin calculus and Stein's techniques. The reader is referred to [20] for the former and [2] for the latter and see [19] for the connection of the two fields. The analysis of this connection is a very active area of research, and our paper illustrates the usefulness of the analysis even for MCMC.

The paper is organized as follows. The numerical simulations are provided in the next section. We also illustrate the limitation of the MpCN algorithm in Section 2.3.4. In Section 3, high-dimensional asymptotic properties will be studied. We will show that the pCN algorithm is worse than the classical RWM algorithm for the class of heavy-tailed target probability distributions. On the other hand, the MpCN algorithm attains a better convergence rate than the RWM algorithm. Proofs are relegated to Section 4. In the appendix, Section A includes a short introduction to the Malliavin calculus and Stein's techniques. Section B provides some properties for consistency of MCMC.

Finally, we note that, in our current study, we only describe the usefulness of our algorithm for the class of heavy-tailed target probability distributions. However, this heavy tail assumption is just an example of target probability distribution that is difficult to approximate by MCMC. Our method works well for a large class of target probability distributions. These include the Bayesian posterior distributions for ergodic/non-ergodic settings of diffusion processes, point processes such as Hawkes process, and other i.i.d. models. Some diffusion results can be found in [14] and [13]. (More precisely, a version of MpCN. See Section 3.4 for the detail.) The target

probability distribution is very complicated although it is not heavy-tailed. The performance of the Gaussian RWM algorithm was quite poor due to the complexity. However, the new method worked well as described in Figure 1 of [14].

1.1. Notation

For a real number x , $[x]$ is the integer part of $x > 0$. The Euclidean space \mathbb{R}^d is equipped with the norm $\|x\| = (\sum_{i=1}^d x_i^2)^{1/2}$ and inner product $\langle x, y \rangle = \sum_{i=1}^d x_i y_i$ for vectors $x = (x_1, \dots, x_d), y = (y_1, \dots, y_d) \in \mathbb{R}^d$. The $d \times d$ -identity matrix is denoted by I_d . In this space, the normal distribution with mean $\mu \in \mathbb{R}^d$ and variance covariance matrix Σ is denoted by $N_d(\mu, \Sigma)$, and its density is denoted by $\phi_d(x; \mu, \Sigma)$. When $d = 1$, we use simpler notation $N(\mu, \sigma^2)$ for the normal distribution with mean μ and variance σ^2 with its density $\phi(x; \mu, \sigma^2)$.

For a real-valued function $f : E \rightarrow \mathbb{R}$, we consider the supremum norm $\|f\|_\infty = \sup_{x \in E} |f(x)|$. For a signed measure ν on (E, \mathcal{E}) , we consider the total variation norm $\|\nu\|_{\text{TV}} = \sup_{A \in \mathcal{E}} |\nu(A)| = 2^{-1} \sup_{\|f\|_\infty \leq 1} |\nu(f)|$ where $\nu(f) = \nu f = \int_E f(x) \nu(dx)$.

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{L}(X)$ is the law of random variable X . For a random variable X with an event $A \in \mathcal{F}$, we write

$$\mathbb{E}[X, A] = \mathbb{E}[X 1_A].$$

The weak convergence of the random variable is denoted by $X_n \Rightarrow X$. When the sequence $\mathcal{L}(X_n)$ ($n = 1, 2, \dots$) is tight, we write $X_n = O_{\mathbb{P}}(1)$, and write $X_n = o_{\mathbb{P}}(1)$ if $X_n \Rightarrow 0$. Write $X|Y$ for the conditional distribution of X given Y .

2. The MpCN algorithm and its performance

2.1. The pCN algorithm

In this section, we describe two Metropolis–Hastings algorithms. Let $P(dx)$ be a probability measure on (E, \mathcal{E}) with the probability density function $p(x)$ with respect to a σ -finite measure $\nu(dx)$. The Metropolis–Hastings algorithm generates a Markov chain $\{X_m\}_m$ with the Markov kernel $K(x, dy)$ on (E, \mathcal{E}) defined by the following: Let $R(x, dy)$ be a Markov kernel with $\mathcal{E}^{\otimes 2}$ -jointly measurable probability density function $r(x, y)$ of $R(x, dy)$ with respect to $\nu(dy)$. Set $X_0 \in E$ and for $m \geq 1$,

$$\begin{cases} X_m^* \sim R(X_{m-1}, dx), \\ X_m = \begin{cases} X_m^* & \text{with probability } \alpha(X_{m-1}, X_m^*), \\ X_{m-1} & \text{with probability } 1 - \alpha(X_{m-1}, X_m^*), \end{cases} \end{cases}$$

where $R(x, dy)$ is called the proposal kernel, and $\alpha(x, y)$ is called the acceptance ratio which is defined by

$$\alpha(x, y) = \min \left\{ 1, \frac{p(y)r(y, x)}{p(x)r(x, y)} \right\}.$$

This acceptance probability is designed to satisfy

$$P(dx)R(x, dy)\alpha(x, y) = P(dy)R(y, dx)\alpha(y, x). \tag{2.1}$$

The Markov chain is called reversible with respect to $P(dx)$ if

$$P(dx)K(x, dy) = P(dy)K(y, dx).$$

If the acceptance ratio satisfies (2.1), then the Markov chain has reversibility. See monograph [23] or review [29] for further details.

When $E = \mathbb{R}^d$, the most popular choice is $R(x, dy) = N_d(x, \Sigma)$ where Σ is a positive definite matrix. However, this popular choice reveals the limitation in high-dimension as described in [24]. Another approach was proposed by [17] which sometimes works better. Let P_d be a probability measure on \mathbb{R}^d with density $p_d(x)$ with respect to the Lebesgue measure. In this paper, the following algorithm that generate a Markov chain $X^d = \{X_m^d\}_{m \in \mathbb{N}_0}$ is called the pre-conditioned Crank–Nicolson (pCN) algorithm for the target probability distribution P_d if $\rho \in (0, 1)$ and X_0^d is a \mathbb{R}^d -valued random variable, and for $m \geq 1$,

$$\begin{cases} X_m^{d*} = \sqrt{\rho}X_{m-1}^d + \sqrt{1-\rho}W_m^d, & W_m^d \sim N_d(0, I_d), \\ X_m^d = \begin{cases} X_m^{d*} & \text{with probability } \alpha_d(X_{m-1}^d, X_m^{d*}), \\ X_{m-1}^d & \text{with probability } 1 - \alpha_d(X_{m-1}^d, X_m^{d*}), \end{cases} \end{cases} \tag{2.2}$$

where $\alpha_d(x, y) = \min\{1, p_d(y)\phi_d(x; 0, I_d)/p_d(x)\phi_d(y; 0, I_d)\}$. In this case, the proposal kernel is $R_d(x, dy) = N_d(\sqrt{\rho}x, (1-\rho)I_d)$. Then

$$\phi_d(x; 0, I_d) dx R_d(x, dy) = \phi_d(y; 0, I_d) dy R_d(y, dx).$$

Thus, the proposal kernel is reversible with respect to the standard normal distribution.

The pCN algorithm works well when the target probability distribution P_d is approximately Gaussian distribution. However, we will see that the algorithm becomes even worse for a class of heavy-tailed target probability distributions.

2.2. The MpCN algorithm

In this paper, we propose the following algorithm that generates a Markov chain $X^d = \{X_m^d\}_{m \in \mathbb{N}_0}$: Set $\rho \in (0, 1)$ and set X_0^d as a \mathbb{R}^d -valued random variable, and for $m \geq 1$, generates independent random variables $W_m^d, \tilde{W}_m^d \sim N_d(0, I_d)$ and set

$$X_m^{d*} = \sqrt{\rho}X_{m-1}^d + \sqrt{1-\rho}\|X_{m-1}^d\| \frac{W_m^d}{\|\tilde{W}_m^d\|}.$$

Then we set

$$X_m^d = \begin{cases} X_m^{d*} & \text{with probability } \alpha_d(X_{m-1}^d, X_m^{d*}), \\ X_{m-1}^d & \text{with probability } 1 - \alpha_d(X_{m-1}^d, X_m^{d*}), \end{cases}$$

where

$$\alpha_d(x, y) = \min\{1, p_d(y)\|x\|^{-d}/p_d(x)\|y\|^{-d}\}.$$

In this paper, this algorithm is called the mixed pre-conditioned Crank–Nicolson (MpCN) algorithm for the target probability distribution P_d .

Since $d^{1/2}W_m^d/\|\tilde{W}_m^d\|$ follows the t -distribution with d -degrees of freedom in \mathbb{R}^d (see (3.3)), the proposal kernel $R_d(x, dy)$ of the MpCN algorithm is expressed by

$$R_d(x, dy) = r_d(x, y) dy := c \left(\frac{1}{\sqrt{1-\rho}\|x\|d^{-1/2}} \right)^d \left[1 + \frac{1}{d} \left\| \frac{y - \sqrt{\rho}x}{\sqrt{1-\rho}\|x\|d^{-1/2}} \right\|^2 \right]^{-d} dy,$$

where $c = \Gamma(d)/\Gamma(d/2)d^{d/2}\pi^{d/2}$ is the normalizing constant. Then, by taking a σ -finite measure $\bar{P}_d(dx) = \bar{p}_d(x) dx := \|x\|^{-d} dx$, we have

$$\bar{P}_d(dx)R_d(x, dy) = c(1-\rho)^{d/2}d^{d/2}[\|x\|^2 + \|y\|^2 - 2\sqrt{\rho}\langle x, y \rangle]^{-d} dx dy = \bar{P}_d(dy)R_d(y, dx).$$

Thus, the proposal kernel is \bar{P}_d -reversible. The expression of the acceptance probability of the MpCN algorithm comes from this property, since

$$\frac{r_d(y, x)}{r_d(x, y)} = \frac{\bar{p}_d(x)}{\bar{p}_d(y)} = \frac{\|x\|^{-d}}{\|y\|^{-d}}.$$

Since \bar{P}_d has a heavier tail than those of Gaussian distributions, we expect that this new method works well even for heavy-tailed target probability distributions. We will now check the performance of simulation.

2.3. Numerical results

We consider two kinds of numerical experiments.

Efficiency of MpCN algorithm: In Sections 2.3.1–2.3.3, we illustrate efficiency of the MpCN algorithm. We run $M = 10^8$ iterations (no burn-in) for each of the following algorithms:

1. The RWM algorithm with Gaussian proposal distribution. More precisely, the update x^* from the current value x is generated by $x^* = x + \sigma_d \varepsilon$ where ε follows the standard normal distribution and $\sigma_d^2 = 1/d$ in this simulation.
2. The RWM algorithm with the t -distribution as the proposal distribution (two degrees of freedom). More precisely, $x^* = x + \sigma_d \varepsilon$ where ε follows the t -distribution with two degrees of freedom and $\sigma_d^2 = 1/d$ in this simulation.
3. The pCN algorithm for $\rho = 0.8$.
4. The MpCN algorithm for $\rho = 0.8$.

The target probability distributions are the following.

- (a) The standard normal distribution.
- (b) The t -distribution (two degrees of freedom).
- (c) A perturbation of the t -distribution.

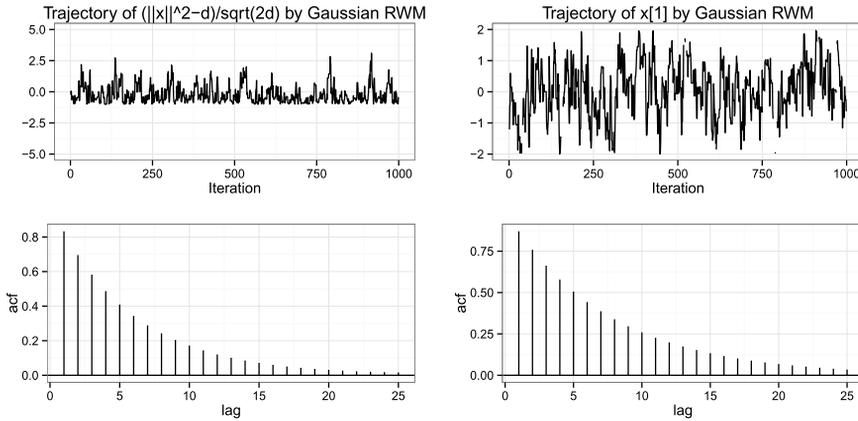


Figure 1. The RWM algorithm with Gaussian proposal distribution for $P_d = N_d(0, I_d)$ for $d = 2$.

For each target probability distribution and each algorithm, we generate a single Markov chain $\{X_m^d\}_m$ with initial value $X_0^d \sim N_d(0, I_d)$ and plot four figures as in Figure 1.

This example is just for an illustration. The target probability distribution is the two dimensional standard normal distribution and the MCMC is the RWM algorithm with Gaussian proposal distribution. These four plots are

- (i) Trajectory of the normalized distance from the origin. When the target probability distribution is the standard normal distribution, we plot $\{(2d)^{-1/2}(\|X_m^d\|^2 - d)\}_m$ and for other cases, we plot $\{\|X_m^d\|^2/d\}_m$ (upper left).
- (ii) The autocorrelation plot of the above (bottom left).
- (iii) Trajectory $\{X_{m,1}^d\}_m$ where $X_m^d = (X_{m,1}^d, \dots, X_{m,d}^d)$ (upper right).
- (iv) The autocorrelation plot of the above (bottom right).

The simulation results are illustrated in Sections 2.3.1–2.3.3.

Shift perturbation effect: We also illustrate the limitation of our algorithm and how to avoid it in Section 2.3.4. The target probability distribution is $P_d(\xi \mathbf{1} - dx)$ where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^d$ and

$$\xi = 0, 1, 2, 3, \quad \text{or} \quad 4$$

and P_d is

- (a) the standard normal distribution, or
- (b) the t -distribution (two degrees of freedom).

We plot

- (ii) the autocorrelation plot of $\{(2d)^{-1/2}(\|X_m^d - \xi \mathbf{1}\|^2 - d)\}_m$ for the standard normal distribution, and plot that of $\{\|X_m^d - \xi \mathbf{1}\|^2/d\}_m$ for the t -distribution for $\xi \in \{0, 1, 2, 3, 4\}$.

Although we can not apply our theoretical results to this non-spherically symmetric target probability distribution, it is a good example to illustrate the limitation of our algorithm. The performance of MCMC for the shift $\xi \mathbf{1}$ will illustrate shift sensitivity of the MCMC algorithms. The RWM algorithms are, essentially, free from the shift. However, the pCN and MpCN are sensitive to this effect. Fortunately, this effect can be avoided by a simple estimate of the center of the target probability distribution. We will show the results with and without this estimation. Therefore, in practice, sometimes the performance can be improved by using the estimates of the center and the covariance structure.

Since RWM algorithm is free from this effect, we only consider the pCN and MpCN algorithms. We can compare the results in this section to that of the RWM algorithms in Sections 2.3.1 and 2.3.2.

2.3.1. The standard normal distribution in \mathbb{R}^{20}

Set $P_d = N_d(0, I_d)$ for $d = 20$. For this case, the convergence rate for the RWM algorithm is d . On the other hand, the pCN and MpCN algorithms attain consistency, and so these algorithms are better than the RWM algorithm (see Section 3.4). The simulation shows that the performance of the RWM algorithm for the Gaussian proposal and the t -distribution proposal are similar (Figures 2 and 3), and that for the pCN and MpCN algorithms are also similar (Figures 4 and 5) and are much better than the former two algorithms.

We also observed the effective sample size (ESS; [6]. See also 12.3.5 of [23]) by using coda package ([22]) in R. The results in Table 1 are calculated from 5000 samples after 5000 burn-in samples for 50 parallel runs and calculated the average over d coordinates. The value of the ESS was multiplied by a factor of 100 to reflect the percentage of the total MCMC iterations that can be considered as independent draws from the posterior. We choose the tuning parameter of the RWM algorithm so that the acceptance probability is around 25%. The results are not surprising; as in autocorrelation plot, the pCN and MpCN algorithms work better than the RWM algorithm.

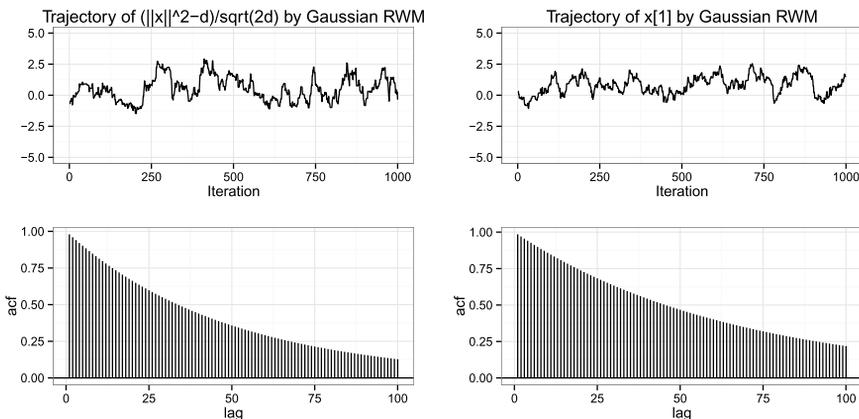


Figure 2. The RWM algorithm with Gaussian proposal distribution for $P_d = N_d(0, I_d)$ for $d = 20$.

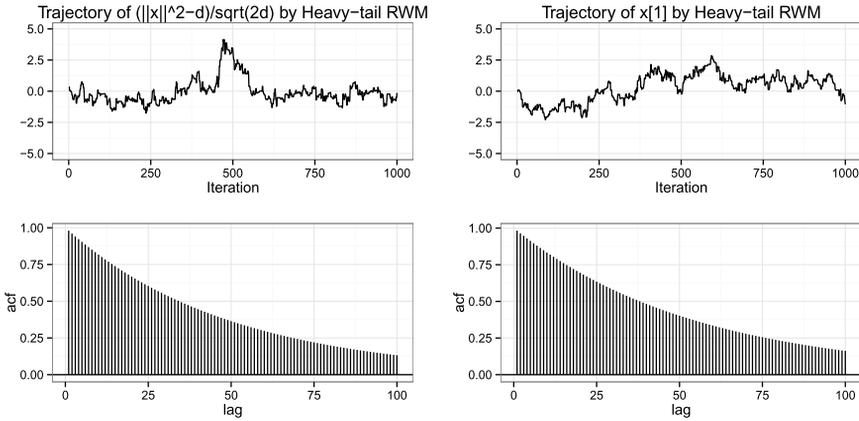


Figure 3. The RWM algorithm with t -distribution as the proposal distribution for $P_d = N_d(0, I_d)$ for $d = 20$.

2.3.2. P_d is the t -distribution with two degrees of freedom in \mathbb{R}^{20}

Set P_d as the t -distribution with $\nu = 2$ degrees of freedom with the scale parameter $\sigma = 5$ and shift $\mu = 0$ for $d = 20$. Recall that the probability distribution function is given by

$$p_d(x) = \frac{\Gamma((\nu + d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}\sigma^d(1 + \|x - \mu\|^2/\nu)^{(\nu+d)/2}}. \tag{2.3}$$

For this case, the convergence rate for the RWM algorithm is d^2 . The pCN algorithm is much worse than the rate, and the MpCN algorithm attains much better rate d (Theorems 3.1 and 3.2).

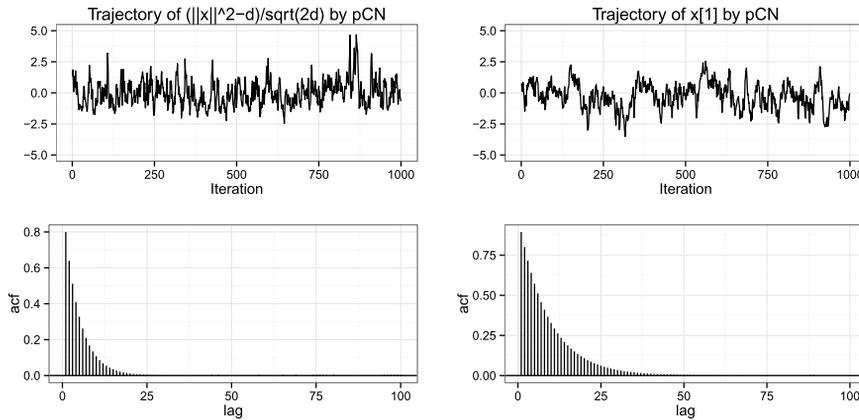


Figure 4. The pCN algorithm for $P_d = N_d(0, I_d)$ for $d = 20$.

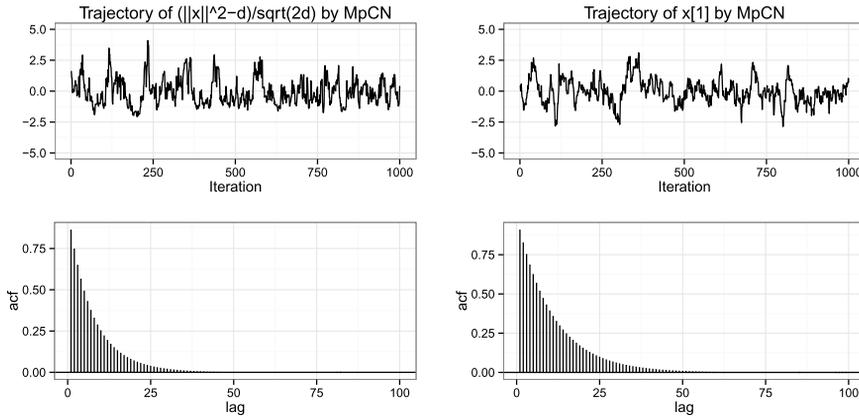


Figure 5. The MpCN algorithm for $P_d = N_d(0, I_d)$ for $d = 20$.

In the simulation, the MpCN algorithm is much better than other algorithms which correspond to the theoretical result (Figures 6–9).

As in Table 2, the estimated value of the ESS also reveals the limitation of the pCN, and efficiency of the MpCN. The estimated ESS for the MpCN algorithm is better than that of the RWM algorithm as in autocorrelation plots. However, we remark that for this high-dimension heavy tail case, the estimate of the ESS may not be reliable.

2.3.3. A perturbation of the t -distribution

We show the performance of the MpCN algorithm when the target probability distribution is not spherically symmetric. Let P_{20} be a probability measure in \mathbb{R}^{20} with the probability density function

$$p_{20}(x_1, x_2, \dots, x_{20}) \propto \left(1 + \sum_{i=1}^{20} \left(\frac{x_i - 1}{5} \right)^2 + |x_1| + \sin(x_2)/2 \right)^{-(4+20)/2} .$$

Table 1. Effective sample size for $P_d = N_d(0, I_d)$ for $d = 20$

Method	Effective sample size	Acceptance probability
RWM	0.828	0.226
RWM t -distribution	0.594	0.254
pCN	2.770	0.980
MpCN	2.375	0.801

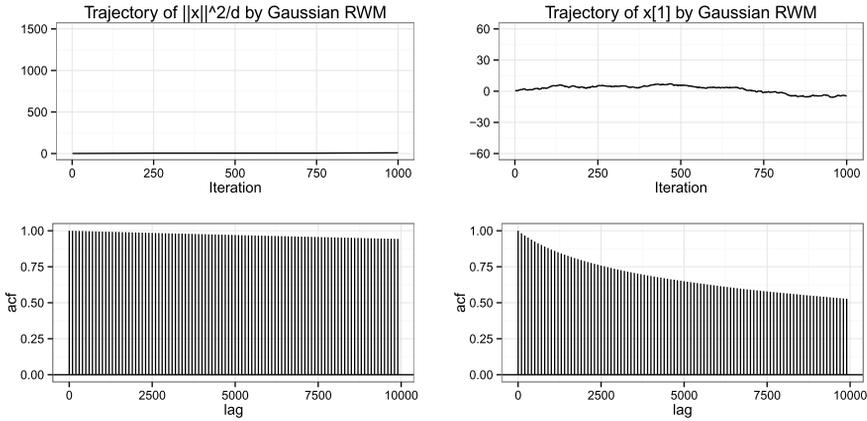


Figure 6. The RWM algorithm with Gaussian proposal distribution when t -distribution is the target probability distribution.

The distribution is not scaled mixture, and so we can not say anything for the convergence rate for this case. However, by simulation, we observe that the MpCN algorithm is much better than other algorithms (Figures 10–13).

Again, we calculated the ESS in Table 3. We can still observe the gap between the MpCN and other algorithms though it is smaller than that of the t -distribution case.

2.3.4. *Shift-perturbation of spherically symmetric target probability distributions*

Let $P_d = N_d(\xi \mathbf{1}, I_d)$, where $\xi = 0, 1, 2, 3, 4$ for $d = 20$ and consider the pCN and MpCN algorithms. Compare the results of the RWM algorithms in Section 2.3.1 (bottom left figures of Fig-

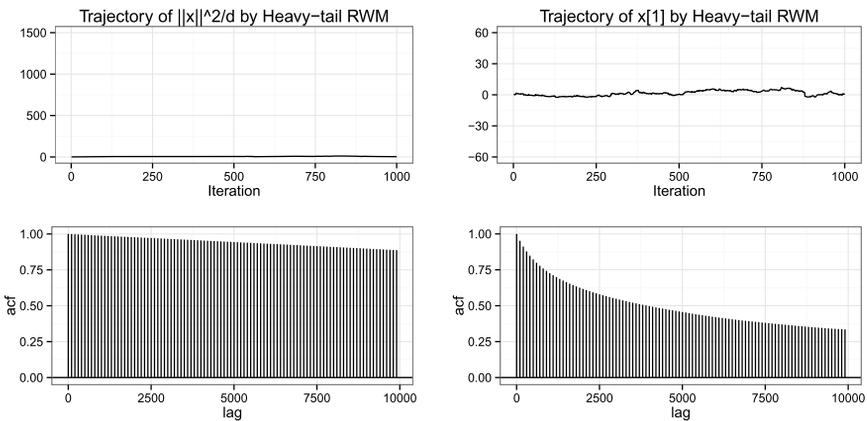


Figure 7. The RWM algorithm with t -distribution as the proposal distribution and the target probability distribution is also the t -distribution.

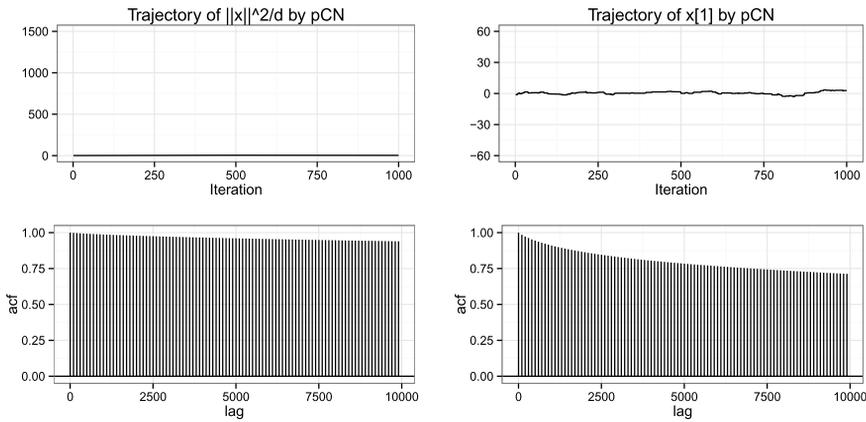


Figure 8. The pCN algorithm when t -distribution is the target probability distribution.

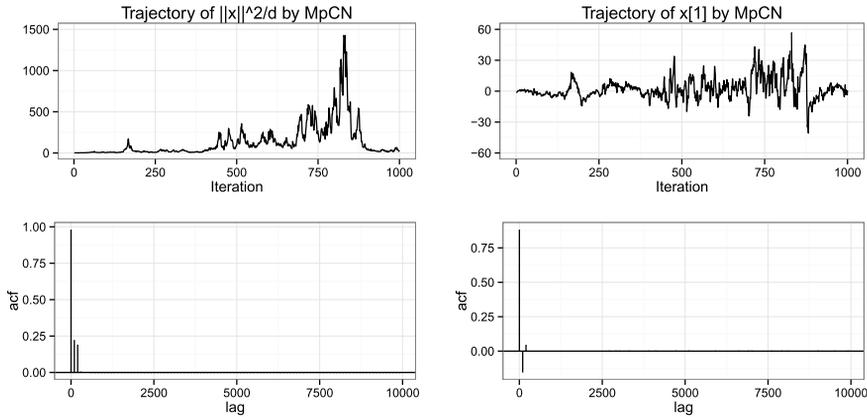


Figure 9. The MpCN algorithm when t -distribution is the target probability distribution.

Table 2. Effective sample size for t -distribution

Method	Effective sample size	Acceptance probability
RWM	0.385	0.194
RWM t -distribution	0.498	0.259
pCN	0.052	0.053
MpCN	3.300	0.941

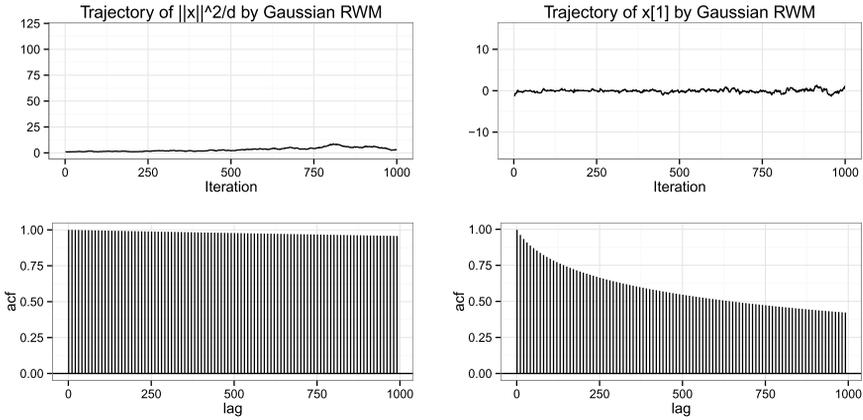


Figure 10. The RWM algorithm with Gaussian proposal distribution when the perturbed t -distribution is the target probability distribution.

ures 2 and 3). Figure 14 illustrates that although the performances of pCN and MpCN algorithms are much better than the RWM algorithms when $\xi = 0$, it is sensitive to the value of ξ . Therefore for the light-tail target probability distribution in high-dimension, when the high-probability region is far from the origin, it is important to shift the target probability distribution in advance. For example, first, calculate rough estimate $\hat{\xi}$ of the center of the target probability distribution $P_d(dx)$, and then run the MCMC algorithm for $P_d(-\hat{\xi} + dx)$. Tempering strategies might be useful for the rough estimate of the center of the target probability distribution as in [14].

Next figure (Figure 15) is a result of the pCN and MpCN algorithm with a simple estimation of the center of the target probability distribution. We run $M = 10^3$ iteration of the pCN or MpCN

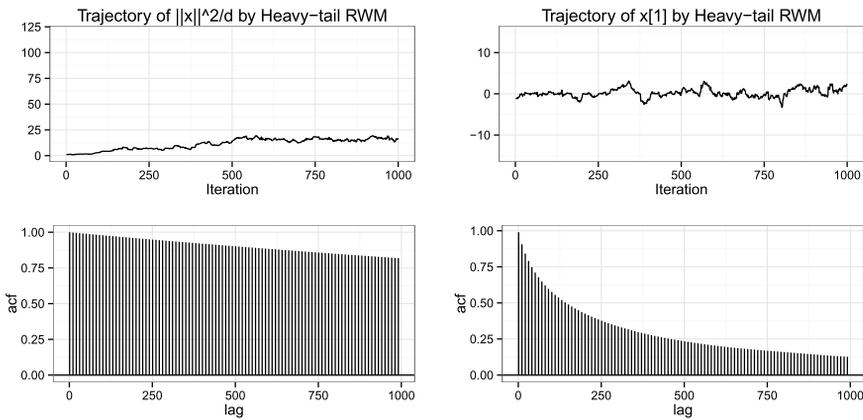


Figure 11. The RWM algorithm with t -distribution as the proposal distribution and the target probability distribution is the perturbed t -distribution.

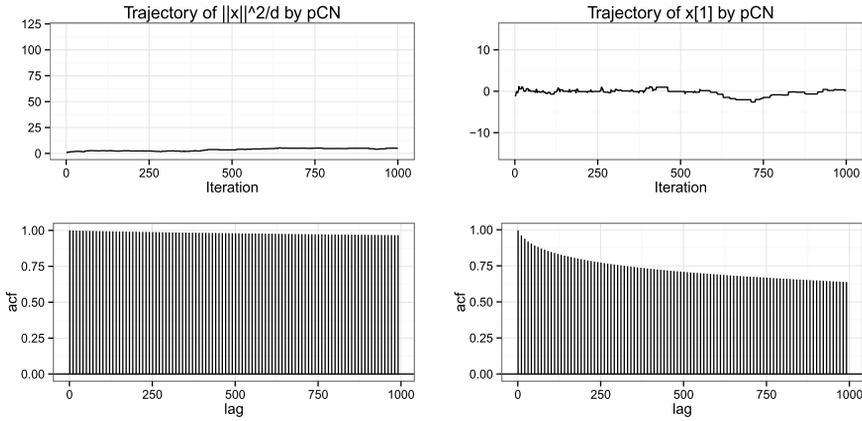


Figure 12. The pCN algorithm when the perturbed t -distribution is the target probability distribution.

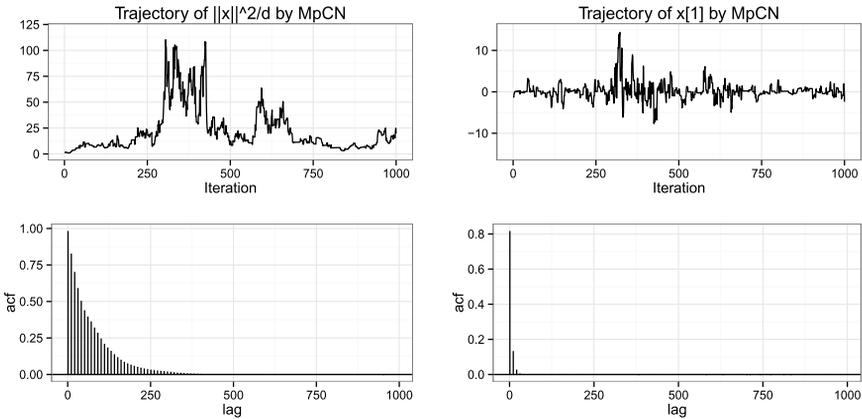


Figure 13. The MpCN algorithm when the perturbed t -distribution is the target probability distribution.

Table 3. Effective sample size for the perturbed t -distribution

Method	Effective sample size	Acceptance probability
RWM	0.549	0.226
RWM t -distribution	0.484	0.195
pCN	0.129	0.088
MpCN	1.863	0.418

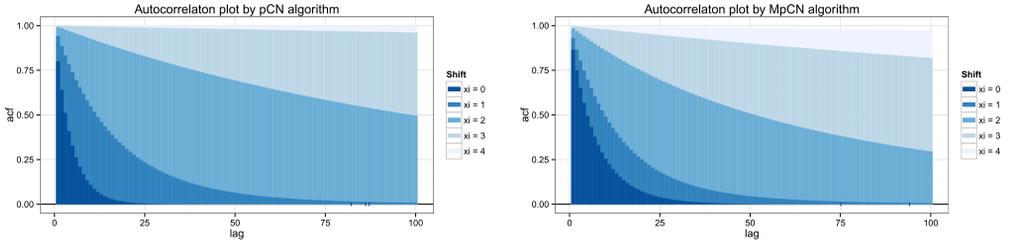


Figure 14. Autocorrelation plots for the pCN and MpCN algorithms for shifted normal distributions.

algorithm to calculate

$$\hat{\xi} = M^{-1} \sum_{m=0}^{M-1} X_m^d \tag{2.4}$$

and then run $M = 10^8$ iteration of the pCN or MpCN algorithm for the target probability distribution $P_d(-\hat{\xi} + dx)$. The effect of the shift is considerably weakened.

We consider the t -distribution (2.3) with $\nu = 2$ and $\sigma = 5$ and $\mu = \xi \mathbf{1}$ where $\xi = 0, 1, 2, 3, 4$ for $d = 20$. Compare it to the results in Section 2.3.2 for the RWM algorithms (bottom left figures of Figures 6 and 7). Compared to the light-tailed distribution, the effect of the shift is small for the MpCN algorithm, and the five autocorrelation plots are overlapped in Figure 16.

The next figure (Figure 17), which is almost identical to the previous one, is a result of $M = 10^8$ iteration of the pCN and MpCN algorithm with a simple estimation of the target probability distribution (2.4) by $M = 10^3$ iteration. Thus for heavy-tailed target probability distribution, the effect of shift perturbation is small, and the gain of the estimation of the center is also small.

3. High-dimensional asymptotic theory

We consider a sequence of the target probability 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

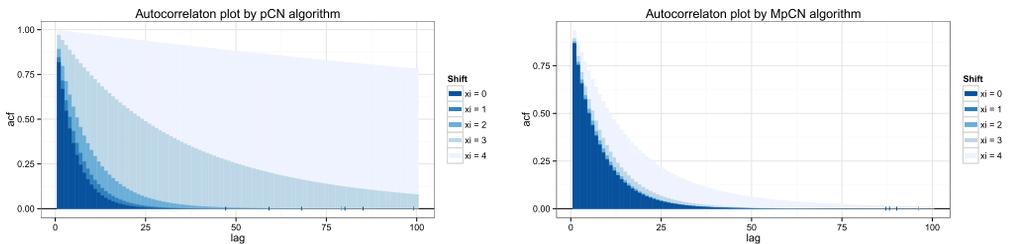


Figure 15. Autocorrelation plots for the pCN and MpCN algorithms for shifted normal distributions with an initial estimate of the center of the target probability distribution.

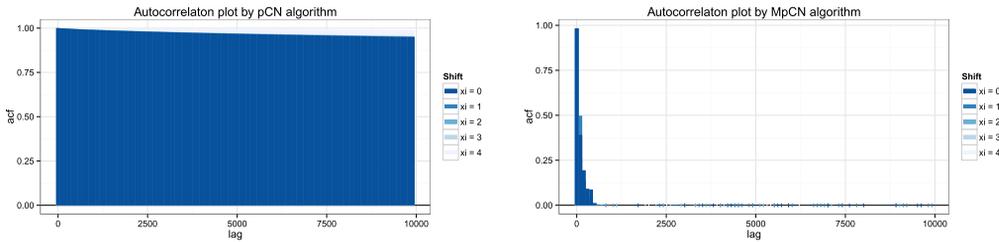


Figure 16. Autocorrelation plots for the pCN and MpCN algorithms for shifted t -distributions.

mixture of the normal distribution. Furthermore, our asymptotic setting is that the number of dimension d goes to infinity while the mixing distribution Q of P_d is unchanged. Note that in our results, stationarity and reversibility of the Markov chains generated by Markov chain Monte Carlo algorithms are essential. However, this can be weakened as explained in Lemma 4 of [11].

3.1. Consistency

In this section, we generalize the consistency of MCMC studied in [11]. For each $d \in \mathbb{N}$, suppose that Markov chain Monte Carlo method \mathcal{M}^d generates a Markov chain $\{X_m^d; m \in \mathbb{N}_0\}$ with the invariant probability distribution P_d . The consistency defined in [11] is the property such that the integral $P_d(f) = \int f(x)P_d(dx)$ we want to calculate is approximated by a Monte Carlo simulated value $\frac{1}{M} \sum_{m=0}^{M-1} f(X_m^d)$ after a reasonable number of iterations M . For example, regular Gibbs sampler should satisfy this type of property (more precisely, local consistency, see [11]) when d is the sample size of the data.

In the current context, the state space for $X^d = \{X_m^d; m \in \mathbb{N}_0\}$ ($d \in \mathbb{N}$) changes as $d \rightarrow \infty$ that is inconvenient for further analysis. As in [24] and [10], to overcome the difficulty, we set a projection $\pi_{d,k}$ ($k \leq d$) to a finite subset by

$$\pi_{d,k}(x) = (x_i)_{i=1,\dots,k} \quad (x = (x_i)_{i=1,\dots,d}).$$

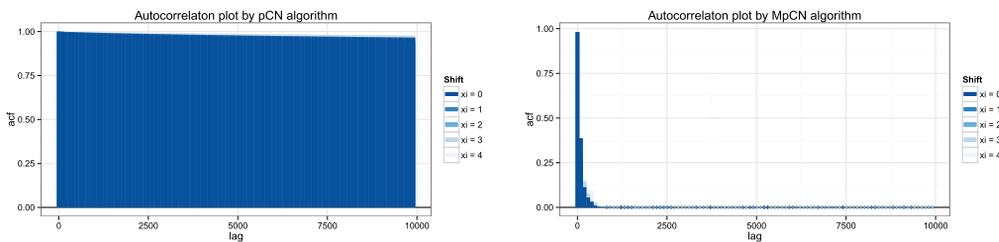


Figure 17. Autocorrelation plots for the pCN and MpCN algorithms for shifted t -distributions with an initial estimate of the center of the probability measure.

Other possibility to overcome this difficulty is to embed \mathbb{R}^d into $\mathbb{R}^{\mathbb{N}}$ under suitable metric as in [1]. We do not follow this approach since it is not obvious how to embed our new algorithm into $\mathbb{R}^{\mathbb{N}}$, and also, we want to avoid further technical difficulties to deal with infinite dimensional convergence of Markov processes.

Definition 3.1 (Consistency for non-fixed dimensional case). A family \mathcal{M}^d ($d \in \mathbb{N}$) is called consistent if for any $k \in \mathbb{N}$, $M_d \rightarrow \infty$ and for any bounded continuous function $f : \mathbb{R}^k \rightarrow \mathbb{R}$,

$$\frac{1}{M_d} \sum_{m=0}^{M_d-1} f \circ \pi_{d,k}(X_m^d) - P_d(f \circ \pi_{d,k}) = o_{\mathbb{P}}(1)(d \rightarrow \infty). \tag{3.1}$$

It is not hard to show that the pCN and MpCN algorithms are consistent for a class of light-tailed target probability distributions. However, when P_d is a heavy-tailed distribution, these methods do not have consistency, but weak consistency defined by the following.

Definition 3.2 (Weak Consistency). A family \mathcal{M}^d ($d \in \mathbb{N}$) is called weakly consistent with rate T_d if (3.1) is satisfied for any $M_d \rightarrow \infty$ such that $M_d/T_d \rightarrow \infty$. We will call the rate T_d , the convergence rate. If $T_d/d^k \rightarrow 0$ for some $k \in \mathbb{N}$, we call that it has a polynomial rate of convergence.

The rate T_d corresponds to the number of iterations until good convergence. Therefore the smaller, the better. Note that if the family \mathcal{M}^d is consistent, then the convergence rate is $T_d = 1$.

Example 3.1. Let \mathcal{M}^d be the RWM algorithm with target probability distribution $P_d = N_d(0, I_d)$ and the proposal distribution $N_d(x, \sigma^2/d)$ for $\sigma > 0$. We assume stationarity, that is, $X_0^d \sim P_d$ for the Markov chain $\{X_m^d\}_m$ generated by \mathcal{M}^d . Then as in Theorem 1.1 of [24], for any $k \in \mathbb{N}$, $Y_t^d := \pi_{d,k}(X_{[dt]}^d)$ converges weakly to a k -dimensional Ornstein–Uhlenbeck process. Then \mathcal{M}^d is weakly consistent and the rate is d by Lemma B.3.

When the target probability distribution is heavy-tailed, the performance of MCMC algorithms is quite different from that for the light-tailed case. In [10], we showed that the convergence rate for the RWM algorithm is d^2 for heavy-tailed target probability distribution. We will show that this rate becomes d for the MpCN algorithm.

3.2. Assumption for the target probability distribution

In this subsection, we describe the class of target probability distributions considered in this paper. We want to study the property of the MpCN algorithm for this class. The analysis becomes much more complicated than that for the light-tailed target probability distribution. To avoid technical difficulties, we want to set the class as minimal as possible. More precisely, we only consider scale mixtures of the normal distribution. The class is not so rich, but it is sufficient for our purpose since it includes many important heavy-tailed target probability distributions such as the t -distribution and the stable distribution.

Let $Q(d\sigma^2)$ be a probability measure on $(0, \infty)$. Let P_d be a scale mixture of the normal distribution defined by

$$P_d = \mathcal{L}(X_0^d), \quad Q_d = \mathcal{L}(\|X_0^d\|^2/d), \tag{3.2}$$

where $X_0^d | \sigma^2 \sim N_d(0, \sigma^2 I_d)$ and $\sigma^2 \sim Q$. Note that $Q_d \rightarrow Q$ as $d \rightarrow \infty$ since $\|X_0^d\|^2/d \rightarrow \sigma^2$ a.s. In this setup, P_d and Q_d have probability density functions p_d and q_d that satisfy

$$p_d(x) \propto \|x\|^{2-d} q_d\left(\frac{\|x\|^2}{d}\right).$$

Assumption 3.1. Probability distribution Q has the strictly positive continuously differentiable probability density function $q(y)$. Each $q(y)$ and $q'(y)$ vanishes at $+0$ and $+\infty$.

Example 3.2 (Student t -distribution). The probability distribution function of the t -distribution with $\nu > 0$ degree of freedom is

$$p_d(x) = \frac{\Gamma((\nu + d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}(1 + \|x\|^2/\nu)^{(\nu+d)/2}}. \tag{3.3}$$

In this case, Q is the inverse chi-squared distribution with ν -degree of freedom with probability distribution function $q(y) \propto y^{-\nu/2-1}e^{-\nu/(2y)}$. It is straightforward to check that $q(y)$ and $q'(y)$ vanishes at $+0$ and $+\infty$. For properties of the (multivariate) t -distribution, see [16].

Example 3.3 (Stable distribution). Let $\alpha \in (0, 2)$. If P_d is the rotationally symmetric α -stable distribution with characteristic function $\int \exp(i\langle t, x \rangle) P_d(dx) = \exp(-\|t\|^2/2\|^{-\alpha/2})$, then Q is $\alpha/2$ -stable distribution on the half line with Laplace transform $\int \exp(-ty) Q(dy) = \exp(-t^{\alpha/2})$ for $t > 0$. Although there is no closed form of probability density function $q(x)$, all derivatives of $q(x)$ are continuous and vanishes at 0 and ∞ . See Section 14 of [26].

For this class of target probability distributions, the acceptance ratio of the MpCN algorithm can be written in the following form:

$$\alpha_d(x, y) = \min\left\{1, \frac{p_d(y)\|x\|^{-d}}{p_d(y)\|y\|^{-d}}\right\} = \min\left\{1, \frac{\tilde{q}_d(r_d(y))}{\tilde{q}_d(r_d(x))}\right\}, \tag{3.4}$$

where

$$\tilde{q}(r) = 2e^{2r}q(e^{2r}), \quad \tilde{q}_d(r) = 2e^{2r}q_d(e^{2r}), \quad r_d(x) = \frac{1}{2} \log(\|x\|^2/d).$$

We write \tilde{Q} and \tilde{Q}_d for probability measure with densities \tilde{q} and \tilde{q}_d . Note that if $y \sim Q(dy)$ and $y_d \sim Q_d(dy_d)$, then $(\log y)/2 \sim \tilde{Q}$ and $(\log y_d)/2 \sim \tilde{Q}_d$. In particular, $\tilde{Q}_d \rightarrow \tilde{Q}$.

3.3. Main results

We describe the main results in this paper. The proofs will be given in Section 4. Set $\rho \in (0, 1)$. We assume that Q is a probability measure on $(0, \infty)$ and for each $d \in \mathbb{N}$, P_d is the scale mixture of the normal distribution defined in (3.2) with the mixing measure Q . We also assume *stationarity* of the process, that is, $X_0^d \sim P_d$. The pCN algorithm does not work well for the class of target probability distribution.

Theorem 3.1. *The pCN algorithm does not have a polynomial rate of convergence if Assumption 3.1 is satisfied.*

On the other hand, the MpCN algorithm still has a good convergence property for this class. Recall that the convergence rate for the RWM algorithm is d^2 as studied in [10]. Let $\eta = \eta(\rho) = \sqrt{1 - \rho}/2$.

Proposition 3.1. *Let Q satisfy Assumption 3.1. Let X^d be a stationary Markov chain generated by the MpCN algorithm and let $Y_t^d = r_d(X_{[dt]}^d)$. Then $Y^d = (Y_t^d)_t$ converges to the stationary ergodic process $Y = (Y_t)_t$ (in Skorohod's topology) that is the solution of*

$$dY_t = a(Y_t) dt + \sqrt{b(Y_t)} dW_t; \quad Y_0 \sim \tilde{Q}, \tag{3.5}$$

where $\{W_t\}_t$ is the standard Brownian motion and

$$a(y) = \frac{\eta}{2} (\log \tilde{q})'(y), \quad b(y) = \eta^2.$$

Theorem 3.2. *Let Q satisfy Assumption 3.1. Then the MpCN algorithm has the convergence rate d .*

3.4. Discussion

In this article, we proposed the MpCN algorithm and provide high-dimensional asymptotic results. The proposal kernel $R_d(x, dy)$ used in MpCN is \tilde{P}_d -reversible where $\tilde{P}_d(dx) = \|x\|^{-d} dx$, and so this is a special case of MCMC that uses reversible proposal kernel. The relation to the target probability distribution P_d and \tilde{P}_d is quite important. If \tilde{P}_d has a heavier-tail than that of P_d , then MCMC behaves relatively well. On the other hand, if \tilde{P}_d has a lighter-tail, it becomes quite poor. The RWM algorithm has $\tilde{P}_d =$ Lebesgue measure. This is a robust choice in the sense that $\sup_{x \in \mathbb{R}^d} dP_d/d\tilde{P}_d$ is bounded for large class of target probability distributions, but it loses efficiency to pay the price as described in [10]. On the other hand, the pCN algorithm, which has $\tilde{P}_d = N_d(0, I_d)$, does not work well except some specific cases. The proposed algorithm, MpCN is in the middle of these algorithms. It is robust and works well.

It is possible to consider a more general class of the MpCN algorithm: Let \bar{Q} be any σ -finite measure on $(0, \infty)$ and set $\bar{P}_d(dx) = \bar{p}_d(x) dx$ where $\bar{p}_d(x) := \int_{z=0}^\infty \phi_d(x; 0, zI_d) \bar{Q}(dz)$. For

$m \geq 1$, set

$$\begin{cases} Z_m^d \sim \phi_d(X_{m-1}^d; 0, zI_d) \overline{Q}(dz) / \overline{p}_d(X_{m-1}^d), \\ X_m^{d*} = \sqrt{\rho} X_{m-1}^d + \sqrt{(1-\rho)} Z_m^d W_m^d, \quad W_m^d \sim N_d(0, I_d), \\ X_m^d = \begin{cases} X_m^{d*} & \text{with probability } \alpha_d(X_{m-1}^d, X_m^{d*}), \\ X_{m-1}^d & \text{with probability } 1 - \alpha_d(X_{m-1}^d, X_m^{d*}), \end{cases} \end{cases}$$

where $\alpha_d(x, y) = \min\{1, p_d(y)\overline{p}_d(x)/p_d(x)\overline{p}_d(y)\}$ assuming that $\overline{p}_d(x) < \infty$ for any $x \in \mathbb{R}^d$. For example, in [14], they set $\overline{Q}(dz) \propto z^{-\nu/2-1} e^{-\nu/(2z)}$. See also [7]. The MpCN algorithm studied in this paper corresponds to $\overline{Q}(dz) = z^{-1} dz$.

In this article, we did not mention ergodic properties. Ergodicity is also studied for this algorithm. It is geometrically ergodic for broader class of target probability distributions than that of the RWM algorithm. In particular, the MpCN algorithm can be geometrically ergodic even for a class of heavy-tailed target probability distributions. See [12].

Also, we did not prove asymptotic properties for a class of light-tailed target probability distributions, such as $P_d = N_d(0, \sigma^2 I_d)$ for $\sigma^2 > 0$. The MpCN algorithm has consistency in this class, but the pCN algorithm has consistency only if $\sigma^2 = 1$. The proof is not difficult but bit complicated which uses different techniques from that used in this paper. Therefore we omit the proof in this paper to focus on the heavy tail case.

Finally, we want to remark that the class of target probability distributions that we consider is fairly restrictive. The extension of the class is not simple and probably requires new techniques. However, as illustrated in the simulation, we believe that by using our restrictive class, we have successfully described the actual behavior of the MCMC algorithms. In practice, as discussed in Section 2.3.3, normalizing the target probability distribution may improves the performance. That is, it might be better to apply the pCN and the MpCN algorithms for the target probability distribution $P_d(dx)$ with scaling $x \mapsto \Sigma^{-1/2}(x - \mu)$, where μ and Σ are estimated values of the center and the covariance (correlation structure) of P_d .

4. Proofs

Let $\rho \in (0, 1)$, $d \in \mathbb{N}$ and let $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ be the usual Euclidean norm and the inner product. For $x \in \mathbb{R}^d \setminus \{0\}$ and for independent random variables $W^d, \tilde{W}^d \sim N_d(0, I_d)$, let

$$F_d(x) = \frac{d^{1/2}}{2\eta} \left\{ \log \left\| \sqrt{\rho}x + \sqrt{1-\rho}\|x\| \frac{W^d}{\|\tilde{W}^d\|} \right\|^2 - \log \|x\|^2 \right\}.$$

This random variable essentially determines the behavior of the asymptotic properties of the Markov chain generated by the MpCN kernel since the law of $\log \|X_m^{d*}\| - \log \|X_{m-1}^d\|$ is the same as that of $d^{-1/2}\eta F_d(X_{m-1}^d)$. We first describe some properties of F_d which is useful to analyze asymptotic properties of the MpCN algorithm and then study some properties of Q_d . The proofs of the main results will be described in Section 4.3.

4.1. Some properties of F_d

To establish asymptotic properties of the MpCN kernel, we need to know some asymptotic behaviors of the random variable $F_d(x)$. We will prove that the law of F_d is very close to $N(0, 1)$ and so the behavior of $\log \|X_m^{d*}\| - \log \|X_{m-1}^d\|$ is similar to the Gaussian random-walk. First, we observe that $F_d(x)$ is symmetric about the origin.

Proposition 4.1. *The law of $F_d(x)$ does not depend on the choice of $x \in \mathbb{R}^d \setminus \{0\}$ and it is symmetric about the origin, that is, $\mathbb{P}(F_d(x) < \eta) = \mathbb{P}(F_d(x) > -\eta)$ for any $\eta \in \mathbb{R}$.*

Proof. Observe that for independent random variables $W_d, \tilde{W}_d \sim N_d(0, I_d)$,

$$\begin{aligned} F_d(x) &= \frac{d^{1/2}}{2\eta} \log \left\| \sqrt{\rho} \frac{x}{\|x\|} + \sqrt{1-\rho} \frac{W^d}{\|\tilde{W}^d\|} \right\|^2 \\ &= \frac{d^{1/2}}{2\eta} \log \left(\rho + 2\sqrt{\rho(1-\rho)} \frac{\|W^d\|}{\|\tilde{W}^d\|} U^d(x) + (1-\rho) \frac{\|W^d\|^2}{\|\tilde{W}^d\|^2} \right), \end{aligned}$$

where

$$U^d(x) = \left\langle \frac{x}{\|x\|}, \frac{W^d}{\|W^d\|} \right\rangle.$$

In order to prove that $\mathcal{L}(F_d(x))$ does not depend on x , we show that $\mathcal{L}(U^d(x))$ has the property. But it is obvious since $\mathcal{L}(U^d(x)) = \mathcal{L}(U^d(\alpha Vx))$ for any unitary matrix V and $\alpha > 0$.

Next, we prove that the law of F_d is symmetric about the origin. Since the law of $F_d(x)$ does not depend on x , it has also the same law as that of

$$F_d := F_d \left(\frac{\tilde{W}^d}{\|\tilde{W}^d\|} \right) = \frac{d^{1/2}}{2\eta} (\log \|\sqrt{\rho}\tilde{W}^d + \sqrt{1-\rho}W^d\|^2 - \log \|\tilde{W}^d\|^2),$$

where we used the fact that the random variables W^d and $\|\tilde{W}^d\|$ in $F_d(x)$ are independent from $\tilde{W}^d/\|\tilde{W}^d\|$. Recall that $(\tilde{W}^d, \sqrt{\rho}\tilde{W}^d + \sqrt{1-\rho}W^d)$ is an exchangeable pair, that is,

$$\mathcal{L}(\tilde{W}^d, \sqrt{\rho}\tilde{W}^d + \sqrt{1-\rho}W^d) = \mathcal{L}(\sqrt{\rho}\tilde{W}^d + \sqrt{1-\rho}W^d, \tilde{W}^d).$$

Thus $\mathcal{L}(F_d) = \mathcal{L}(-F_d)$, that is, the law of F_d is symmetric about the origin. □

By the above result, without loss of generality, we can set

$$F_d := \frac{d^{1/2}}{2\eta} (\log \|\sqrt{\rho}\tilde{W}^d + \sqrt{1-\rho}W^d\|^2 - \log \|\tilde{W}^d\|^2), \tag{4.1}$$

for independent random variables $W_d, \tilde{W}_d \sim N_d(0, I_d)$, since the law does not change. The other properties of F_d can be studied via the Malliavin calculus in Section A. The results are summarized as follows.

Proposition 4.2. *The random variable F_d has a density with respect to the Lebesgue measure and the following properties are satisfied.*

1. $\sup_{d \in \mathbb{N}} \mathbb{E}[|F_d|^4] < \infty$.
2. *There exists a constant $C > 0$ such that for any absolutely continuous function $f : \mathbb{R} \rightarrow \mathbb{R}$,*

$$|\mathbb{E}[F_d f(F_d)] - \mathbb{E}[f'(F_d)]| \leq C d^{-1/2} \|f'\|_\infty.$$

3. $\|\mathcal{L}(F_d) - N(0, 1)\|_{TV} \rightarrow 0$.

4.2. Some properties of Q_d

We need the following technical result for the chi-squared distribution.

Lemma 4.1. *For $d \in \mathbb{N}$, let ξ_d follows the chi-squared distribution with d degrees of freedom. For $k \in \mathbb{Z}$ there exists $C > 0$ such that for any $d \in \mathbb{N}$ with $-2k < d$,*

$$\left| \mathbb{E} \left[\left(\frac{\xi_d}{d} \right)^k \right] - 1 \right| \leq \frac{C}{d}.$$

Moreover, for $k > 2$,

$$\mathbb{E} \left[\left\{ d^{1/2} \left(\frac{\xi_d}{d} - 1 \right) \right\}^k \right]^{1/k} \leq (k-1)\sqrt{2}.$$

Proof. The first clam follows from

$$\mathbb{E} \left[\left(\frac{\xi_d}{d} \right)^k \right] = \frac{2^k \Gamma(k + d/2)}{d^k \Gamma(d/2)} = \begin{cases} \left(1 + \frac{2}{d}\right) \cdots \left(1 + \frac{2}{d}(k-1)\right) & \text{if } k > 0, \\ \left(1 - \frac{2}{d}\right) \cdots \left(1 - \frac{2}{d}k\right) & \text{if } k < 0 \end{cases}$$

The second part follows from Example A.2. □

An immediate corollary from the lemma is $\mathbb{E}[(d/\xi_d - 1)^2] = o(1)$ since

$$\mathbb{E} \left[\left(\frac{d}{\xi_d} - 1 \right)^2 \right] = \mathbb{E} \left[\left(\frac{d}{\xi_d} \right)^2 - 1 \right] - 2\mathbb{E} \left[\left(\frac{d}{\xi_d} \right) - 1 \right] = O(d^{-1}) = o(1). \tag{4.2}$$

Proposition 4.3. $\|q_d - q\|_\infty \rightarrow 0, \|q'_d - q'\|_\infty \rightarrow 0$.

Proof. The probability density q_d is

$$q_d(x) = \int_0^\infty g \left(y; \frac{d}{2}, \frac{d}{2} \right) q \left(\frac{x}{y} \right) \frac{dy}{y},$$

where $g(y; \nu, \alpha)$ is the probability density function of the Gamma distribution defined in Introduction. Therefore, in expectation form,

$$q_d(x) = \mathbb{E} \left[q \left(x \frac{d}{\xi_d} \right) \frac{d}{\xi_d} \right],$$

where ξ_d follows the chi-squared distribution with d -degrees of freedom. Then

$$\begin{aligned} |q_d(x) - q(x)| &= \left| \mathbb{E} \left[q \left(x \frac{d}{\xi_d} \right) \frac{d}{\xi_d} \right] - q(x) \right| \\ &\leq \|q\|_\infty \mathbb{E} \left[\left| \frac{d}{\xi_d} - 1 \right| \right] + \mathbb{E} \left[\left| q \left(x \frac{d}{\xi_d} \right) - q(x) \right| \right]. \end{aligned}$$

The first term in the right-hand side is $o(1)$ by (4.2). For the second term, by uniform continuity of $q(x)$ together with $\lim_{x \rightarrow \infty} q(x) = 0$, we can find $\delta > 0$ and $C > 0$ for any $\varepsilon > 0$ such that $|x - y| \leq \delta$ implies $|q(x) - q(y)| < \varepsilon$ and $x \geq C$ implies $q(x) < \varepsilon$. Then, for $x \leq 2C$,

$$\begin{aligned} \mathbb{E} \left[\left| q \left(x \frac{d}{\xi_d} \right) - q(x) \right| \right] &\leq \mathbb{E} \left[\left| q \left(x \frac{d}{\xi_d} \right) - q(x) \right|, \left| \frac{d}{\xi_d} - 1 \right| > \frac{\delta}{2C} \right] \\ &\quad + \mathbb{E} \left[\left| q \left(x \frac{d}{\xi_d} \right) - q(x) \right|, \left| \frac{d}{\xi_d} - 1 \right| \leq \frac{\delta}{2C} \right] \\ &\leq \|q\|_\infty \left(\frac{\delta}{2C} \right)^{-2} \mathbb{E} \left[\left| \frac{d}{\xi_d} - 1 \right|^2 \right] + \varepsilon \rightarrow \varepsilon \quad (d \rightarrow \infty) \end{aligned}$$

and for $x > 2C$,

$$\begin{aligned} \mathbb{E} \left[\left| q \left(x \frac{d}{\xi_d} \right) - q(x) \right| \right] &\leq \mathbb{E} \left[\left| q \left(x \frac{d}{\xi_d} \right) - q(x) \right|, \left| \frac{d}{\xi_d} - 1 \right| > 1/2 \right] \\ &\quad + \mathbb{E} \left[\left| q \left(x \frac{d}{\xi_d} \right) - q(x) \right|, \left| \frac{d}{\xi_d} - 1 \right| \leq 1/2 \right] \\ &\leq \|q\|_\infty 4 \mathbb{E} \left[\left| \frac{d}{\xi_d} - 1 \right|^2 \right] + \varepsilon \rightarrow \varepsilon \quad (d \rightarrow \infty). \end{aligned}$$

Thus $\|q_\infty - q\|_\infty \rightarrow 0$. The proof is completely the same for q'_d . □

By this property, \tilde{q}_d and \tilde{q}'_d converges to \tilde{q} and \tilde{q}' uniformly on any compact set.

4.3. Proof of Proposition 3.1

We prove convergence of Markov chain to the diffusion process. For this purpose, we embed Markov chain to a continuous Markov process. Let $r_d(x) = \frac{1}{2} \log \left(\frac{\|x\|^2}{d} \right)$ ($x \in \mathbb{R}^d \setminus \{0\}$) and write

$$R_m^d = r_d(X_m^d), \quad R_m^{d*} = r_d(X_m^{d*}) \tag{4.3}$$

Let N^d be a Poisson process which is independent of $\{X_m^d, X_m^{d*}\}_m$. We will assume that N^d has the intensity dt , that is, $\mathbb{E}[N_t^d] = dt$. Set $\tilde{Y}_t^d = R_{N_t^d}^d$. Then the process \tilde{Y}^d is a pure step Markov process with generator

$$Af(y) = d^{-1}(\mathbb{E}[f(R_1^d)|R_0^d = y] - f(y)).$$

See Section 4.2 of [5] for the detail. We will apply Theorem IX.4.21 of [9] to the process \tilde{Y}^d . If \tilde{Y}^d converges to a limit, then Y^d defined in Proposition 3.1 converges to the same limit by Lemma B.1.

Proof of Proposition 3.1. For the proof, we need to show some asymptotic properties of conditional distribution of R_1^d given $R_0^d = y \in \mathbb{R}$. For notational simplicity, we write y for R_0^d , and write \mathbb{P}_y and \mathbb{E}_y for the conditional probability and expectation given $R_0^d = y$. By using this notation, we have $R_1^{d*} = y + \eta d^{-1/2} F_d$ regarding W_1^d and \tilde{W}_1^d as W^d and \tilde{W}^d . Let

$$\begin{cases} a_d(y) = d\mathbb{E}_y[R_1^d - R_0^d], \\ b_d(y) = d\mathbb{E}_y[(R_1^d - R_0^d)^2], \\ c_d(y) = d\mathbb{E}_y[(R_1^d - R_0^d)^4]. \end{cases} \tag{4.4}$$

First, we consider estimate of $a_d(y)$. We have

$$a_d(y) = d\mathbb{E}_y\left[(R_1^{d*} - R_0^d) \min\left\{1, \frac{\tilde{q}_d(R_1^{d*})}{\tilde{q}_d(R_0^d)}\right\}\right] = \eta d^{1/2}\mathbb{E}_y\left[F_d \min\left\{1, \frac{\tilde{q}_d(y + \eta d^{-1/2} F_d)}{\tilde{q}_d(y)}\right\}\right].$$

Since \tilde{q}_d may not be bounded, we introduce a localization function $\psi_\varepsilon : \mathbb{R} \rightarrow [0, 1]$ which is a C^∞ function and satisfies $\psi_\varepsilon(x) = 1$ if $|x| \leq \varepsilon$ and $\psi_\varepsilon(x) = 0$ when $|x| > 2\varepsilon$ for $\varepsilon > 0$. Then $\psi_\varepsilon(d^{-1/2} F_d) = 1 + o_{\mathbb{P}}(1)$. Moreover,

$$\mathbb{E}[|F_d| |1 - \psi_\varepsilon(d^{-1/2} F_d)|] \leq \mathbb{E}[|F_d|, |d^{-1/2} F_d| > \varepsilon] \leq d^{-3/2} \mathbb{E}[|F_d|^4] / \varepsilon^3 = O(d^{-3/2}),$$

by Markov's inequality. Suppose that $\tilde{q}'(y) > 0$. We can find an open bounded neighborhood \mathcal{O} of y such that $\inf_{z \in \mathcal{O}} \tilde{q}'(z) > \delta$ for some $\delta > 0$. In addition, since \tilde{q}'_d converges to \tilde{q}' uniformly on a bounded set \mathcal{O} , we have $\inf_{z \in \mathcal{O}} \tilde{q}'_d(z) > \delta/2$ sufficiently large d , and so for $z \in \mathcal{O}$,

$$\min\left\{1, \frac{\tilde{q}_d(z)}{\tilde{q}_d(y)}\right\} = \begin{cases} 1 & \text{if } z > y, \\ \frac{\tilde{q}_d(z)}{\tilde{q}_d(y)} & \text{if } z \leq y \end{cases} = \frac{\tilde{q}_d(\min\{z, y\})}{\tilde{q}_d(y)}.$$

Choose ε so that $(y - 2\varepsilon, y + 2\varepsilon) \subset \mathcal{O}$. Then by Proposition 4.2

$$\begin{aligned} a_d(y) &= \eta d^{1/2} \mathbb{E}_y \left[F_d \frac{\tilde{q}_d(y + \eta d^{-1/2} \min\{F_d, 0\})}{\tilde{q}_d(y)} \psi_\varepsilon(d^{-1/2} F_d) \right] + o(1) \\ &= \eta^2 \mathbb{E}_y \left[\frac{\tilde{q}'_d(y + \eta d^{-1/2} F_d)}{\tilde{q}_d(y)} \psi_\varepsilon(d^{-1/2} F_d), F_d \leq 0 \right] + o(1), \end{aligned}$$

where we note that contribution from the term which includes ψ'_ε is $o(1)$, and the convergence is uniform on \mathcal{O} . Then, since \tilde{q}_d and \tilde{q}'_d converges to \tilde{q} and \tilde{q}' uniformly on \mathcal{O} , and $\psi_\varepsilon(d^{-1/2}F_d) \rightarrow 1$ in probability, we have

$$a_d(y) = \eta^2 \frac{\tilde{q}'(y)}{\tilde{q}(y)} \mathbb{P}_y(F_d \leq 0) + o(1) = \frac{\eta^2}{2} \frac{\tilde{q}'(y)}{\tilde{q}(y)} + o(1),$$

where we used the fact that $\mathcal{L}(F_d) \rightarrow N(0, 1)$ in the last equation. We omit the proof of the case $\tilde{q}'(y) < 0$ since the argument is the same.

Suppose that $\tilde{q}'(y) = 0$. Choose a bounded neighborhood \mathcal{O} of y so that $\inf_{y \in \mathcal{O}} \tilde{q}(y) > \delta$ for some $\delta > 0$, and so $\inf_{y \in \mathcal{O}} \tilde{q}_d(y) > \delta/2$ sufficiently large d . For any $\varepsilon > 0$ we can find a neighborhood $\mathcal{O}' \subset \mathcal{O}$ such that $|\tilde{q}'(z)| \leq \varepsilon\delta/4$, and hence $|\tilde{q}'_d(z)| \leq \varepsilon\delta/2$ for $z \in \mathcal{O}'$ when d is large enough. Then

$$\left| \min \left\{ 1, \frac{\tilde{q}_d(y + \eta d^{-1/2} F_d)}{\tilde{q}_d(y)} \right\} - 1 \right| \leq \frac{\sup_{z \in \mathcal{O}'} |\tilde{q}'_d(z)|}{\tilde{q}_d(y)} \eta d^{-1/2} |F_d| \leq \varepsilon \eta d^{-1/2} |F_d|,$$

and hence

$$|a_d(y)| \leq \varepsilon \eta^2 \mathbb{E}_y[|F_d|^2] \leq \varepsilon \eta^2 \mathbb{E}_y[|F_d|^4]^{1/2}.$$

Since we can choose any $\varepsilon > 0$, we have $a_d(y) \rightarrow 0$ locally uniformly in \mathcal{O} . Thus, we proved that

$$a_d(y) \rightarrow \frac{\eta^2}{2} \frac{\tilde{q}'(y)}{\tilde{q}(y)}$$

locally uniformly.

Next, we prove the convergence of b_d and c_d . Observe

$$b_d(y) = d \mathbb{E}_y[(R_1^d - R_0^d)^2] = \eta^2 \mathbb{E}_y \left[F_d^2 \min \left\{ 1, \frac{\tilde{q}_d(y + \eta d^{-1/2} F_d)}{\tilde{q}_d(y)} \right\} \right].$$

Then,

$$\begin{aligned} b_d(y) &= \eta^2 \mathbb{E}_y \left[F_d^2 \min \left\{ 1, \frac{\tilde{q}_d(y + \eta d^{-1/2} F_d)}{\tilde{q}_d(y)} \right\}, |F_d| \leq d^{1/4} \right] + o(1) \\ &= \eta^2 \mathbb{E}_y [F_d^2, |F_d| \leq d^{1/4}] + o(1) \\ &= \eta^2 \mathbb{E}_y [F_d^2] + o(1) \rightarrow \eta^2 + o(1), \end{aligned}$$

where we used Markov's inequality twice, and the moment convergence $\mathbb{E}[F_d^2] \rightarrow 1$ comes from convergence in distribution with $\sup_d \mathbb{E}[F_d^4] < \infty$. In the same way,

$$c_d(y) = d \mathbb{E}_y [(R_1^d - R_0^d)^4] \leq \eta^4 d^{-1} \mathbb{E}_y [|F_d|^4] = o(1) \quad (d \rightarrow \infty).$$

Thus, we obtain the convergence of the triplet (4.4). This convergence corresponds to the conditions (i) and (ii) of Theorem IX.4.21 of [9], and the condition (iii) corresponds to $\mathcal{L}(R_0^d) = \tilde{Q}_d \rightarrow \tilde{Q}$, which is obvious. In addition, the existence and uniqueness of the stochastic differential equation (3.5) can be checked for example by Proposition 5.5.22 of [15] and hence condition IX.4.3 (i) holds, and measurability follows by Exercise 6.7.4 of [28] and hence condition IX.4.3 (ii) holds. Thus, the convergence $\tilde{Y}^d \Rightarrow Y$ follows from Theorem IX.4.21 of [9]. Hence Y^d converges to the same limit by Lemma B.1.

Stationarity and ergodicity of Y is yet to be proved. However, stationarity comes from the fact that each Y^d are stationary, and ergodicity comes from positive recurrence by Proposition 5.5.22 of [15]. Hence, the claim follows. \square

Proof of Theorem 3.2. First, we note that all proposed values of the MpCN algorithm are accepted for a finite number of iterations $M \in \mathbb{N}$ in probability 1 since

$$\begin{aligned} \mathbb{P}(X_{m-1}^d = X_m^d \exists m \in \{1, \dots, M - 1\}) &\leq M\mathbb{P}(X_0^d = X_1^d) \\ &= M \left(1 - \mathbb{E} \left[\min \left\{ 1, \frac{\tilde{q}_d(R_0^d + \eta d^{-1/2} F_d)}{\tilde{q}_d(R_0^d)} \right\} \right] \right) \rightarrow 0. \end{aligned}$$

Second, for a finite number of iterations, R_m^d is almost constant, and $\|\tilde{W}_m^d\|^2/d - 1$ is almost 0, that is,

$$\mathbb{P}(|R_{m-1}^d - R_m^d| \geq \varepsilon \exists m \in \{1, \dots, M - 1\}) \leq M\mathbb{P}(|\eta d^{-1/2} F_d| \geq \varepsilon) \rightarrow 0$$

and

$$\mathbb{P}(\|\tilde{W}_m^d\|^2/d - 1 \geq \varepsilon \exists m \in \{1, \dots, M\}) \leq M\mathbb{P}(\|\tilde{W}_1^d\|^2/d - 1 \geq \varepsilon) \rightarrow 0.$$

Let $S_m^d = \pi_{d,k}(X_m^d)$. By above properties of R_m^d and \tilde{W}_m^d , it is not difficult to check that the joint process $\{(R_m^d, S_m^d)\}_m$ converges weakly to $\{(R_m, S_m)\}_m$ defined by

$$\begin{cases} R_m = R_0, \\ S_m = \sqrt{\rho} S_{m-1} + \sqrt{1 - \rho} \exp(R_0) W_m, \end{cases} \quad W_m \sim N_k(0, I_k)$$

for $m \geq 1$, where $R_0 \sim \tilde{Q}$ and $S_0 \sim N_k(0, \exp(2R_0)I_k)$. By Proposition 3.1, the process $Y^d = \{R_{[dt]}^d\}_t$ converges to a stationary ergodic process. Hence, the claim follows by Lemma B.3. \square

4.4. Inconsistency for the pCN algorithm

Lemma 4.2. *Let $\{X_m^d\}_m$ be the Markov chain generated by the pCN algorithm. Then, for any $\varepsilon > 0, k \in \mathbb{N}$,*

$$d^k \mathbb{P}(|R_0^d| > \varepsilon, X_1^d \neq X_0^d) = o(1).$$

Proof. We have

$$\begin{aligned} R_1^{d*} - R_0^d &= \frac{1}{2} \log\left(\frac{\|X_1^{d*}\|^2}{d}\right) - \frac{1}{2} \log\left(\frac{\|X_0^d\|^2}{d}\right) \\ &= \frac{1}{2} \log\left(\rho + \frac{2\sqrt{\rho(1-\rho)}}{\|X_0^d\|} \left\langle \frac{X_0^d}{\|X_0^d\|}, W_1^d \right\rangle + (1-\rho) \frac{\|W_1^d\|^2}{\|X_0^d\|^2}\right). \end{aligned}$$

Let

$$A_d = \left\langle \frac{X_0^d}{\|X_0^d\|}, W_1^d \right\rangle, \quad B_d = d^{-1/2} \frac{\|W_1^d\|^2 - d}{2}.$$

Remark here that $\mathbb{E}[A_d] = \mathbb{E}[B_d] = 0$ and $\sup_d \mathbb{E}[|A_d|^k] < \infty$ and $\sup_d \mathbb{E}[|B_d|^k] < \infty$ for any $k \in \mathbb{N}$ by $A_d \sim N(0, 1)$ and the second part of Lemma 4.1.

Suppose now that $R_0^d > \varepsilon$. Then $e^{2\varepsilon} \leq \|X_0^d\|^2/d$ and

$$\begin{aligned} R_1^{d*} - R_0^d &\leq \frac{1}{2} \log(\rho + 2\sqrt{\rho(1-\rho)}e^{-\varepsilon}|d^{-1/2}A_d| + (1-\rho)e^{-2\varepsilon}(1 + 2d^{-1/2}B_d)) \\ &=: \frac{1}{2} \log(1 - \xi + d^{-1/2}C_d), \end{aligned}$$

where $\xi = 1 - \rho - (1 - \rho)e^{-2\varepsilon} > 0$ and $C_d = c_1|A_d| + c_2B_d$ for some $c_1, c_2 > 0$. Then $\sup_d \mathbb{E}[|C_d|^k] < \infty$. By this fact,

$$\mathbb{P}(R_1^d > R_0^d > \varepsilon) \leq \mathbb{P}(R_0^d > \varepsilon, R_1^{d*} > R_0^d) \leq \mathbb{P}(d^{-1/2}C_d > \xi) \leq d^{-k} \mathbb{E}\left[\left\{\frac{C_d}{\xi}\right\}^{2k}\right] = O(d^{-k})$$

for any $k \in \mathbb{N}$. By the same argument, we can prove

$$\mathbb{P}(R_1^d < R_0^d < -\varepsilon) = O(d^{-k}).$$

Since the Metropolis–Hastings algorithm generates reversible Markov chain, and we assumed stationarity in this paper, (R_0^d, R_1^d) is an exchangeable pair. Thus

$$\mathbb{P}(R_0^d > R_1^d > \varepsilon) = O(d^{-k}), \quad \mathbb{P}(R_0^d < R_1^d < -\varepsilon) = O(d^{-k})$$

and hence

$$\mathbb{P}(|R_0^d| > \varepsilon, R_1^d \neq R_0^d) = O(d^{-k}).$$

On the other hand, since $X_1^d \neq X_0^d$ implies that X_1^{d*} is accepted, hence $R_1^d = R_1^{d*}$ if $X_1^d \neq X_0^d$. Thus

$$\mathbb{P}(R_1^d = R_0^d, X_1^d \neq X_0^d) \leq \mathbb{P}(R_1^{d*} = R_0^d) = \mathbb{P}(F_d = 0) = 0,$$

since $F_d = d^{1/2}(R_1^{d*} - R_0^d)/\eta$ has a probability density function. By using these estimates, we have

$$\begin{aligned} \mathbb{P}(|R_0^d| > \varepsilon, X_1^d \neq X_0^d) &\leq \mathbb{P}(|R_0^d| > \varepsilon, X_1^d \neq X_0^d, R_1^d \neq R_0^d) + \mathbb{P}(|R_0^d| > \varepsilon, X_1^d \neq X_0^d, R_1^d = R_0^d) \\ &\leq \mathbb{P}(|R_0^d| > \varepsilon, R_1^d \neq R_0^d) + \mathbb{P}(X_1^d \neq X_0^d, R_1^d = R_0^d) = O(d^{-k}). \quad \square \end{aligned}$$

Lemma 4.3. *Let P_d be a scale mixture of the Gaussian distribution. Then the pCN algorithm does not have any polynomial rate of convergence if $Q(\{1\}) < 1$.*

Proof. Since $\tilde{Q}(\{0\}) = Q(\{1\}) < 1$, there exists an open set \mathcal{O} which does not include the origin, and $\tilde{Q}(\mathcal{O}) \geq \delta$ for some $\delta > 0$. By $\tilde{Q}_d \rightarrow \tilde{Q}$ and by Lemma 4.2, for any $p \in \mathbb{N}$,

$$\begin{aligned} \liminf_{d \rightarrow \infty} \mathbb{P}(\forall i, j < d^p, X_i^d = X_j^d) &\geq \liminf_{d \rightarrow \infty} \mathbb{P}(R_0^d \in \mathcal{O}, \forall i, j < d^p, X_i^d = X_j^d) \\ &\geq \liminf_{d \rightarrow \infty} \mathbb{P}(R_0^d \in \mathcal{O}) - \limsup_{d \rightarrow \infty} \mathbb{P}(R_0^d \in \mathcal{O}, \exists i, j < d^p, X_i^d \neq X_j^d) \\ &\geq \tilde{Q}(\mathcal{O}) - \limsup_{d \rightarrow \infty} d^p \mathbb{P}(R_0^d \in \mathcal{O}, X_1^d \neq X_0^d) = \tilde{Q}(\mathcal{O}) \geq \delta. \end{aligned}$$

Thus we have the following degenerate property:

$$\liminf_{d \rightarrow \infty} \mathbb{P}\left(\frac{1}{d^p} \sum_{m=0}^{d^p-1} f \circ \pi_{1,d}(X_m^d) = f \circ \pi_{1,d}(X_0^d)\right) \geq \delta$$

for any bounded continuous function $f(x)$.

We show that if this degeneracy holds, we can not have a polynomial rate of consistency. Assume by the way of contradiction that it is weakly consistent with rate T_d where $T_d/d^p \rightarrow 0$. Then the following should also be satisfied:

$$\frac{1}{d^p} \sum_{m=0}^{d^p-1} f \circ \pi_{1,d}(X_m^d) - P_d(f \circ \pi_{1,d}) = o_{\mathbb{P}}(1).$$

Note here that since P_d is the scale mixture of the normal distribution, $\mathcal{L}(\pi_{1,d}(X_0^d)) = P_1$. Hence, we have

$$P_1(\{x; |f(x) - P_1(f)| < \varepsilon\}) = \liminf_{d \rightarrow \infty} \mathbb{P}(|f \circ \pi_{1,d}(X_0^d) - P_d(f \circ \pi_{1,d})| < \varepsilon) \geq \delta$$

for any $\varepsilon > 0$. By monotone convergence theorem, this is possible only if $P_1(\{x; f(x) = c\}) \geq \delta$ for some $c \in \mathbb{R}$, and thus it is not satisfied for example, for $f(x) = \arctan(x)$ since P_1 has a probability density function. Therefore, the pCN algorithm cannot be weakly consistent with rate T_d where $T_d/d^p \rightarrow 0$ for any $p > 0$ and hence the pCN algorithm cannot have a polynomial rate of convergence. □

Appendix A: Properties of F_d via the Malliavin calculus

A.1. Basic operators in the Malliavin calculus

We will study asymptotic properties of F_d defined in (4.1). The basic tool will be the Malliavin calculus. The following is a quick review of the Malliavin calculus. For the detail, see monographs such as [20] and [19].

Abstract Wiener space Let \mathfrak{H} be a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathfrak{H}}$ and the norm $\|h\|_{\mathfrak{H}}^2 = \langle h, h \rangle_{\mathfrak{H}}$. Let $\{W(h); h \in \mathfrak{H}\}$ be an isonormal Gaussian process on $(\Omega, \mathcal{F}, \mathbb{P})$, that is, $W(h)$ is centered Gaussian and $\mathbb{E}[W(g)W(h)] = \langle g, h \rangle_{\mathfrak{H}}$. By this definition, $W(ag + bh) = aW(g) + bW(h)$ a.s. for $a, b \in \mathbb{R}$ and $g, h \in \mathfrak{H}$ since $\mathbb{E}[|W(ag + bh) - (aW(g) + bW(h))|^2] = 0$. We assume that σ -algebra \mathcal{F} is generated by W . This triplet $(W, \mathfrak{H}, \mathbb{P})$ is called an abstract Wiener space.

Wiener-Chaos decomposition Let $L^2(\Omega)$ be the space of square integrable random variables. Let $H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} / n!$ be the n th Hermite polynomial:

$$H_1(x) = x, \quad H_2(x) = \frac{x^2 - 1}{2}, \quad H_3(x) = \frac{x^3 - 3x}{3!}, \dots$$

The Hermite polynomial satisfies $H'_n = H_{n-1}$. By using this fact together with the integration by parts formula, we have $\mathbb{E}[H_n(W(h))] = 0$ and $\mathbb{V}[H_n(W(h))] = 1/n!$ for $\|h\|_{\mathfrak{H}} = 1$. Random variables $H_n(W(h))$ and $H_m(W(h))$ are orthogonal in the sense that

$$\mathbb{E}[H_n(W(h))H_m(W(h))] = 0$$

for $n \neq m$. Write \mathcal{H}_n for the closed linear subspace of $L^2(\Omega)$ generated by a subset $\{H_n(W(h)); h \in \mathfrak{H}, \|h\|_{\mathfrak{H}} = 1\}$. The linear space \mathcal{H}_n is called the n th Wiener chaos. Then Wiener chaoses spans $L^2(\Omega)$: any element $F \in L^2(\Omega)$ can be described by $F = \mathbb{E}[F] + \sum_{n=1}^{\infty} F_n$ for $F_n \in \mathcal{H}_n$, that is, $L^2(\Omega) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n$, where \mathcal{H}_0 is the set of constants. This is called the Wiener-Chaos decomposition or the Wiener-Itô decomposition.

Fréchet derivative A smooth random variables is a random variable with the form $F = f(W(h_1), \dots, W(h_n))$ where $h_i \in \mathfrak{H}$ and f is a C^∞ function such that all derivatives have polynomial growth. Then the Fréchet derivative of F is defined by

$$DF = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(W(h_1), \dots, W(h_n))h_i$$

and so DF is a random variable with values in \mathfrak{H} . For example, $DH_n(W(h)) = H_{n-1}(W(h))h$. We set

$$\|F\|_{\mathbb{D}^{1,2}} := (\mathbb{E}[|F|^2] + \mathbb{E}[\|DF\|_{\mathfrak{H}}^2])^{1/2}.$$

Write $\mathbb{D}^{1,2}$ for the closure of the space of smooth random variables with respect to the norm $\|\cdot\|_{\mathbb{D}^{1,2}}$ and extend D to $\mathbb{D}^{1,2}$. Then for any $F \in L^2(\Omega)$, $\mathbb{E}[\|DF\|_{\mathfrak{H}}^2] = \sum \mathbb{E}[\|DF_n\|_{\mathfrak{H}}^2] < \infty$ if and only if $F \in \mathbb{D}^{1,2}$.

Ornstein–Uhlenbeck semigroup The Ornstein–Uhlenbeck semigroup $(P_t)_{t \geq 0}$ is defined by

$$P_t F = \mathbb{E}[F] + \sum_{n=1}^{\infty} e^{-nt} F_n$$

for $F = \mathbb{E}[F] + \sum_{n=1}^{\infty} F_n$ ($F_n \in \mathcal{H}_n$). The operator L and L^{-1} is defined by

$$LF = \sum_{n=1}^{\infty} -nF_n, \quad L^{-1}F = \sum_{n=1}^{\infty} -F_n/n,$$

where LF can be defined if $\sum n^2 \mathbb{E}[|F_n|^2] < \infty$. Note that we have $\mathbb{E}[\|DL^{-1}F\|_{\mathfrak{H}}^2] = \sum \mathbb{E}[\|DF_n\|_{\mathfrak{H}}^2]/n^2 \leq \mathbb{E}[\|DF\|_{\mathfrak{H}}^2]$.

By the so-called hypercontractivity property of Ornstein–Uhlenbeck operator, we have the following for Wiener chaoses. See Corollary 2.8.14 of [19] for the proof.

Proposition A.1. *Let $F \in \mathcal{H}_n$ for $n \geq 1$. Then for $p > 2$,*

$$\mathbb{E}[|F|^p]^{1/p} \leq (p - 1)^{n/2} \mathbb{E}[|F|^2]^{1/2}.$$

Example A.1. $\mathbb{E}[|H_n(W(h))|^p]^{1/p} \leq (p - 1)^{n/2} / \sqrt{n!}$ for $n \geq 1, p > 2$ and $\|h\|_{\mathfrak{H}} = 1$.

Example A.2. If ξ_d follows the chi-squared distribution with d degrees of freedom, $\mathbb{E}[(\xi_d/d - 1)^p]^{1/p} \leq (p - 1)\sqrt{2/d}$ for $p > 2$. To see this, let $\{e_i\}_{i \in \mathbb{N}}$ be an orthonormal basis of \mathfrak{H} . Then $A_d := \sum_{i=1}^d (W_i(e_i)^2 - 1) \in \mathcal{H}_2$ and it has the same law as that of $\xi_d - d$. Thus we can apply above proposition to $F = A_d/d$.

A.2. Useful bounds from Stein’s method

By integration-by parts formula, we have

$$\mathbb{E}[Ff(F)] = E[f'(F)] \tag{A.1}$$

for $F \sim N(0, 1)$ if f is smooth enough. In fact, the above *Stein’s equation* characterize the standard normal distribution: $F \sim N(0, 1)$ if and only if the above equation is satisfied for a class of smooth functions f . Moreover, the deviation from Stein’s equation bounds the distance between $\mathcal{L}(F)$ and $N(0, 1)$.

Theorem A.1 (Theorem 3.3.1 of [19]).

$$\|\mathcal{L}(F) - N(0, 1)\|_{\text{TV}} \leq \sup_{f \in \mathcal{F}_{\text{TV}}} |\mathbb{E}f'(F) - \mathbb{E}[Ff(F)]|,$$

where $\mathcal{F}_{\text{TV}} = \{f; \|f\|_{\infty} < \sqrt{\pi/2}, \|f'\|_{\infty} \leq 2\}$.

Thus the deviation from normality is bounded by the deviation from Stein’s equation. On the other hand, the deviation from Stein’s equation can be obtained via the Malliavin calculus. The connection between Stein’s technique and the Malliavin calculus is a hot topic in probability and statistics community. The following is the key result for our paper. See Theorem 2.9.1 of [19] for the proof. See also the proof of Theorem 3.1 of [18] to replace smoothness of f by the existence of the density of F .

Proposition A.2 (Theorem 2.9.1 of [19]). *Suppose that $F \in \mathbb{D}^{1,2}$ has a density with respect to the Lebesgue measure. Then for any absolutely continuous function f ,*

$$|\mathbb{E}[(F - \mathbb{E}[F])f(F)] - \mathbb{E}[f'(F)]| \leq \|f'\|_\infty \mathbb{E}[|1 - \langle DF, -DL^{-1}F \rangle_{\mathfrak{H}}|].$$

A.3. Properties of F_d as a random variable in Wiener chaos

We introduce an abstract Wiener space to the MpCN algorithm. Let $\{e_i; i \in \mathbb{Z}\}$ be the orthonormal basis of \mathfrak{H} and set

$$W_1^d = \sum_{i=1}^d W(e_i)e_i, \quad \tilde{W}_1^d = \sum_{i=1}^d W(e_{-i})e_i. \tag{A.2}$$

Then

$$A_d := (\|\sqrt{\rho}\tilde{W}_1^d + \sqrt{1-\rho}W_1^d\|_{\mathfrak{H}}^2 - d)/2, \quad B_d := (\|\tilde{W}_1^d\|_{\mathfrak{H}}^2 - d)/2 \tag{A.3}$$

are in the second Wiener chaos \mathcal{H}_2 since

$$A_d = \sum_{i=1}^d H_2(W(\sqrt{\rho}e_{-i} + \sqrt{1-\rho}e_i)), \quad B_d = \sum_{i=1}^d H_2(W(e_{-i})).$$

We also define

$$C_d = \sum_{i=1}^d W(\sqrt{\rho}e_{-i} + \sqrt{1-\rho}e_i)W(e_{-i}) - d\sqrt{\rho}, \tag{A.4}$$

which is in \mathcal{H}_2 since

$$W(g)W(h) - \langle g, h \rangle_{\mathfrak{H}} = \frac{\|g+h\|_{\mathfrak{H}}^2}{2} H_2\left(W\left(\frac{g+h}{\|g+h\|_{\mathfrak{H}}}\right)\right) - \frac{\|g-h\|_{\mathfrak{H}}^2}{2} H_2\left(W\left(\frac{g-h}{\|g-h\|_{\mathfrak{H}}}\right)\right).$$

Since $H'_n = H_{n-1}$, and $H_1(x) = x$, we have

$$DA_d = \sum_{i=1}^d W(\sqrt{\rho}e_{-i} + \sqrt{1-\rho}e_i)(\sqrt{\rho}e_{-i} + \sqrt{1-\rho}e_i), \quad DB_d = \sum_{i=1}^d W(e_{-i})e_{-i}.$$

By this expression, the joint distribution of (A_d, DA_d) is same as that of (B_d, DB_d) . In addition, $\langle DA_d, DB_d \rangle_{\mathcal{S}} = \sqrt{\rho}C_d + d\rho$. We can interpret F_d defined in (4.1) as a random element in this abstract Wiener space via

$$\begin{aligned} F_d &= \frac{d^{1/2}}{2\eta} (\log(d + 2A_d) - \log(d + 2B_d)) \\ &= \frac{d^{1/2}}{2\eta} \left(\log\left(1 + \frac{2A_d}{d}\right) - \log\left(1 + \frac{2B_d}{d}\right) \right). \end{aligned} \quad (\text{A.5})$$

Note that since $d + 2A_d$ (and hence $d + 2B_d$) follows chi-squared distribution with d degrees of freedom,

$$\mathbb{E} \left[\left(1 + \frac{2A_d}{d} \right)^k \right] = 1 + O(d^{-1}) \quad (\text{A.6})$$

for any $k \in \mathbb{Z}$ by Lemma 4.1. In addition, $\mathbb{E}[|A_d|^k]^{1/k} = O(d^{1/2})$ and $\mathbb{E}[|C_d|^k]^{1/k} = O(d^{1/2})$ for $k \in \mathbb{N}$ by Proposition A.1.

Lemma A.1.

$$\sup_{d \in \mathbb{N}} \mathbb{E}[|F_d|^4] < \infty.$$

Proof. Since the law of F_d is symmetric about the origin, we have

$$\mathbb{E}[F_d^4] = 2\mathbb{E}[(F_d^+)^4],$$

where $x^+ = \max\{0, x\}$. By $\log x \leq x - 1$,

$$\begin{aligned} \mathbb{E}[F_d^4] &= \frac{d^2}{8\eta^4} \mathbb{E}[\{(\log(d + 2A_d) - \log(d + 2B_d))^+\}^4] \\ &\leq \frac{d^2}{8\eta^4} \mathbb{E} \left[\left(\frac{d + 2A_d}{d + 2B_d} - 1 \right)^4 \right]. \end{aligned}$$

By using A_d and B_d defined in (A.3), the right-hand side is

$$d^{-2} \frac{2}{\eta^4} \mathbb{E} \left[\left(\frac{d}{d + 2B_d} \right)^4 (A_d - B_d)^4 \right] \leq d^{-2} \frac{2}{\eta^4} \mathbb{E} \left[\left(\frac{d}{d + 2B_d} \right)^8 \right]^{1/2} \mathbb{E}[(A_d - B_d)^8]^{1/2}$$

by Schwarz inequality. The first expectation in the right-hand side is $O(1)$ by Lemma 4.1. For the second expectation, by Proposition A.1 and Minkowski inequality,

$$\mathbb{E}[(A_d - B_d)^8]^{1/2} \leq (\mathbb{E}[A_d^8]^{1/8} + \mathbb{E}[B_d^8]^{1/8})^4 = O(d^2).$$

Thus we have $\mathbb{E}[F_d^4] = O(1)$. □

We are now going to prove the main result in this section.

Proposition A.3. *There exists $C > 0$ so that for any absolutely continuous function $f : \mathbb{R} \rightarrow \mathbb{R}$,*

$$|\mathbb{E}[F_d f(F_d)] - \mathbb{E}[f'(F_d)]| \leq C d^{-1/2} \|f'\|_\infty.$$

In particular, $\|\mathcal{L}(F_d) - N(0, 1)\|_{\text{TV}} \leq 2C d^{-1/2}$.

Proof. Let f be an absolutely continuous function, and let $C > 0$ be a generic constant which does not depend on the choice of f . Let

$$\xi(F) = Ff(F) - f'(F).$$

Without loss of generality, we may assume $f(0) = 0$ since $\mathbb{E}[F_d] = 0$ by Proposition 4.1. Then

$$|\xi(F)| = |F(f(F) - f(0)) - f'(F)| \leq \|f'\|_\infty (F^2 + 1). \tag{A.7}$$

The first step is to prove

$$|\mathbb{E}[\xi(F_d) - \xi(F_{d,0})]| \leq C d^{-1/2} \|f'\|_\infty \tag{A.8}$$

for a good approximation $F_{d,0} \in \mathbb{D}^{1,2}$, which will be defined below. Recall that F_d can be expressed by (A.5). We replace $\log x$ by a strictly increasing twice continuously differentiable function $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}$ such that $\psi(0) > -\infty$ and $\psi(x) = \log x$ for $x \geq 1/2$. Set

$$F_{d,0} = \frac{d^{1/2}}{2\eta} \left(\psi \left(1 + \frac{2A_d}{d} \right) - \psi \left(1 + \frac{2B_d}{d} \right) \right).$$

Then $F_{d,0} \in \mathbb{D}^{1,2}$ by Proposition 1.2.3 of [20]. Since A_d and B_d have the same law, $\mathbb{E}[F_{d,0}] = 0$. Moreover, $\sup_d \mathbb{E}[F_{d,0}^4] < \infty$ since

$$\begin{aligned} \mathbb{E}[F_{d,0}^4]^{1/4} &\leq \frac{d^{1/2}}{2\eta} \|\psi'\|_\infty \mathbb{E} \left[\left| \frac{2A_d}{d} - \frac{2B_d}{d} \right|^4 \right]^{1/4} \\ &\leq \frac{d^{-1/2}}{\eta} \|\psi'\|_\infty (\mathbb{E}[|A_d|^4]^{1/4} + \mathbb{E}[|B_d|^4]^{1/4}) \leq C, \end{aligned}$$

since $\mathbb{E}[A_d^k]^{1/k} = \mathbb{E}[B_d^k]^{1/k} = O(d^{1/2})$. Set an event

$$E_d = \{F_d \neq F_{d,0}\} \subset \{A_d < -d/4\} \cup \{B_d < -d/4\},$$

which is a rare event since by Markov's inequality,

$$\mathbb{P}(E_d) \leq 2\mathbb{P}(A_d < -d/4) \leq 2\mathbb{E} \left[\left\{ \frac{A_d}{d/4} \right\}^2 \right] = O(d^{-1}).$$

Then by (A.7),

$$\begin{aligned}
|\mathbb{E}[\xi(F_d) - \xi(F_{d,0})]| &= |\mathbb{E}[\xi(F_d) - \xi(F_{d,0}), E_d]| \\
&\leq \mathbb{E}[|\xi(F_d)|, E_d] + \mathbb{E}[|\xi(F_{d,0})|, E_d] \\
&\leq \|f'\|_\infty \{\mathbb{E}[F_d^2 + 1, E_d] + \mathbb{E}[F_{d,0}^2 + 1, E_d]\} \\
&\leq \|f'\|_\infty \{\mathbb{E}[(F_d^2 + 1)^2]^{1/2} + \mathbb{E}[(F_{d,0}^2 + 1)^2]^{1/2}\} \mathbb{P}(E_d)^{1/2},
\end{aligned}$$

where we used the Schwarz inequality in the last inequality. Since the fourth moments of F_d and $F_{d,0}$ are bounded, we obtain (A.8), and the first claim follows.

The second step is to show that $F_{d,0}$ has a density with respect to the Lebesgue measure. With the fact, we will apply Proposition A.2 to $F_{d,0}$ and obtain

$$|\mathbb{E}[\xi(F_{d,0})]| \leq \|f'\|_\infty \mathbb{E}[|1 - \langle DF_{d,0}, -DL^{-1}F_{d,0} \rangle_{\mathfrak{H}}|].$$

In general, a random variable in a finite sum of Wiener chaoses has a density by Theorems 5.1 and 5.2 of [27]. Hence the joint distribution of (A_d, B_d) has a density since the pair is in \mathcal{H}_2 . Since the Jacobian of a map

$$(a, b) \mapsto ((d^{1/2}/2\eta)(\psi(1 + 2a/d) - \psi(1 + 2b/d)), b)$$

is non-degenerate, $F_{d,0}$ has a density.

The third step is to prove

$$|\mathbb{E}[\langle DF_{d,0}, -DL^{-1}F_{d,0} \rangle_{\mathfrak{H}}] - \langle DF_{d,1}, -DL^{-1}F_{d,1} \rangle_{\mathfrak{H}}]| \leq Cd^{-1/2} \quad (\text{A.9})$$

for

$$F_{d,1} = \frac{d^{-1/2}}{\eta}(A_d - B_d) \in \mathcal{H}_2.$$

By the triangular inequality and Hölder's inequality, the left-hand side of (A.9) is bounded above by

$$\begin{aligned}
&\mathbb{E}[|\langle D(F_{d,0} - F_{d,1}), -DL^{-1}F_{d,0} \rangle_{\mathfrak{H}}|] + \mathbb{E}[|\langle DF_{d,1}, -DL^{-1}(F_{d,0} - F_{d,1}) \rangle_{\mathfrak{H}}|] \\
&\leq \mathbb{E}[\|DF_{d,0} - DF_{d,1}\|_{\mathfrak{H}}^2]^{1/2} (\mathbb{E}[\|DF_{d,0}\|_{\mathfrak{H}}^2]^{1/2} + \mathbb{E}[\|DF_{d,1}\|_{\mathfrak{H}}^2]^{1/2}) \\
&\leq \mathbb{E}[\|DF_{d,0} - DF_{d,1}\|_{\mathfrak{H}}^2]^{1/2} (\mathbb{E}[\|DF_{d,0} - DF_{d,1}\|_{\mathfrak{H}}^2]^{1/2} + 2\mathbb{E}[\|DF_{d,1}\|_{\mathfrak{H}}^2]^{1/2}),
\end{aligned}$$

where in the second inequality, we used $\mathbb{E}[\|DL^{-1}F\|_{\mathfrak{H}}^2] \leq \mathbb{E}[\|DF\|_{\mathfrak{H}}^2]$. Thus, the third step will be completed if

$$\mathbb{E}[\|DF_{d,1}\|_{\mathfrak{H}}^2] = O(1), \quad \mathbb{E}[\|DF_{d,0} - DF_{d,1}\|_{\mathfrak{H}}^2] = O(d^{-1/2}). \quad (\text{A.10})$$

For the first part of (A.10), since $DF_{d,1} = (d^{-1/2}/2\eta)(DA_d - DB_d)$, we have

$$\begin{aligned} \|DF_{d,1}\|_{\mathfrak{S}}^2 &= \frac{d^{-1}}{4\eta^2} (\|DA_d\|_{\mathfrak{S}}^2 - 2\langle DA_d, DB_d \rangle_{\mathfrak{S}} + \|DB_d\|_{\mathfrak{S}}^2) \\ &= \frac{d^{-1}}{4\eta^2} ((d + 2A_d) - 2(\sqrt{\rho}C_d + d\rho) + (d + 2B_d)), \end{aligned}$$

where $C_d \in \mathcal{H}_2$ is as in (A.4). Therefore

$$\mathbb{E}[\|DF_{d,1}\|_{\mathfrak{S}}^2] = \frac{d^{-1}}{4\eta^2} (2d - 2d\rho) = O(1).$$

Since (A_d, DA_d) and (B_d, DB_d) have the same law, the left-hand side of the second part of (A.10) is

$$\begin{aligned} &\frac{d^{-1/2}}{\eta} \mathbb{E} \left[\left\| DA_d \left(\psi' \left(1 + \frac{2A_d}{d} \right) - 1 \right) - DB_d \left(\psi' \left(1 + \frac{2B_d}{d} \right) - 1 \right) \right\|_{\mathfrak{S}}^2 \right]^{1/2} \\ &\leq \frac{2d^{-1/2}}{\eta} \mathbb{E} \left[\|DA_d\|_{\mathfrak{S}}^2 \left(\psi' \left(1 + \frac{2A_d}{d} \right) - 1 \right)^2 \right]^{1/2} \\ &\leq \frac{2d^{-1/2}}{\eta} \mathbb{E}[\|DA_d\|_{\mathfrak{S}}^4]^{1/4} \mathbb{E} \left[\left(\psi' \left(1 + \frac{2A_d}{d} \right) - 1 \right)^4 \right]^{1/4}. \end{aligned}$$

We have $\mathbb{E}[\|DA_d\|_{\mathfrak{S}}^4] = \mathbb{E}[(d + 2A_d)^2] = O(d^2)$, and

$$\mathbb{E} \left[\left(\psi' \left(1 + \frac{2A_d}{d} \right) - 1 \right)^4 \right] \leq \|\psi''\|_{\infty} \mathbb{E} \left[\left(\frac{2A_d}{d} \right)^4 \right] = O(d^{-2}).$$

Hence, (A.10) follows.

The fourth step is to show

$$\mathbb{E}[|1 - \langle DF_{d,1}, -DL^{-1}F_{d,1} \rangle_{\mathfrak{S}}|] = O(d^{-1/2}).$$

Since $F_{d,1} \in \mathcal{H}_2$, we have $-L^{-1}F_{d,1} = F_{d,1}/2$, and hence $-DL^{-1}F_{d,1} = DF_{d,1}/2$. Therefore,

$$\langle DF_{d,1}, -DL^{-1}F_{d,1} \rangle_{\mathfrak{S}} = \frac{1}{2} \|DF_{d,1}\|_{\mathfrak{S}}^2$$

and hence together with the fact that $\eta^2 = (1 - \rho)/4$,

$$\begin{aligned} \mathbb{E}[|1 - \langle DF_{d,1}, -DL^{-1}F_{d,1} \rangle_{\mathfrak{S}}|] &= \mathbb{E} \left[\left| 1 - \frac{d^{-1}}{8\eta^2} ((d + 2A_d) - 2(\sqrt{\rho}C_d + d\rho) + (d + 2B_d)) \right| \right] \\ &= \frac{d^{-1}}{8\eta^2} \mathbb{E}[|2A_d - 2\sqrt{\rho}C_d + 2B_d|] = O(d^{-1/2}), \end{aligned}$$

since the second moments of A_d, B_d and C_d are $O(d)$.

By the above four steps, the proof is completed since

$$\begin{aligned} |\mathbb{E}[\hat{\xi}(F_d)]| &\leq |\mathbb{E}[\xi(F_{d,0})]| + Cd^{-1/2}\|f'\|_\infty \\ &\leq \|f'\|_\infty |\mathbb{E}[1 - \langle DF_{d,0}, -DL^{-1}F_{d,0} \rangle_{S_0}]| + Cd^{-1/2}\|f'\|_\infty \\ &\leq \|f'\|_\infty |\mathbb{E}[1 - \langle DF_{d,1}, -DL^{-1}F_{d,1} \rangle_{S_1}]| + Cd^{-1/2}\|f'\|_\infty \\ &\leq Cd^{-1/2}\|f'\|_\infty. \end{aligned}$$

□

Appendix B: Other technical results

B.1. Remark on the time change

Let $\{Z_m^d\}_m$ be a Markov chain, and let N^d be a Poisson process with intensity dt . Let

$$Y_t^d = Z_{[dt]}^d, \quad \tilde{Y}_t^d = Z_{N_t^d}^d.$$

Lemma B.1. *If \tilde{Y}^d converges in law to a process Y , then Y^d converges to the same limit.*

Proof. Write $N^d(t)$ for N_t^d . Let

$$\tau^d(t) = \inf\{s \geq 0; N^d(s) \geq [dt]\}.$$

Then $N^d(\tau^d(t)) = [dt]$ and hence $Y_t^d = \tilde{Y}_{\tau^d}^d$. We apply Proposition VI.6.37 of [9] for \tilde{Y}^d as X^n and Y^d as Y^n in the proposition. It is sufficient to show

$$\lim_{d \rightarrow \infty} \mathbb{P}\left(\sup_{s \leq S} |\tau_s^d - s| > \varepsilon\right) = 0 \tag{B.1}$$

for $\varepsilon > 0, S > 0$. Observe that if $\tau_s^d - s > \varepsilon$, then

$$N^d(s + \varepsilon) \leq N^d(\tau_s^d) = [ds] \leq ds.$$

Similarly, if $\tau_s^d - s < -\varepsilon$ and if $d^{-1} \leq \varepsilon/2$, then

$$N^d(s - \varepsilon) \geq N^d(\tau_s^d) = [ds] \geq ds - 1 \geq d(s - \varepsilon/2).$$

Therefore, for $s \leq S$, we have

$$\begin{aligned} \{|\tau_s^d - s| > \varepsilon\} &\subset \left\{ \frac{N^d(s + \varepsilon)}{d} - (s + \varepsilon) \leq -\varepsilon \right\} \cup \left\{ \frac{N^d((s - \varepsilon)^+)}{d} - (s - \varepsilon)^+ \geq \varepsilon/2 \right\} \\ &\subset \left\{ \sup_{s \leq S'} \left| \frac{N_s^d}{d} - s \right| \geq \frac{\varepsilon}{2} \right\}, \end{aligned}$$

where $S' = S + \varepsilon$. The probability in the left-hand side of (B.1) is bounded by

$$\mathbb{P}\left(\sup_{s \leq S'} \left| \frac{N_s^d}{d} - s \right| \geq \frac{\varepsilon}{2}\right) \leq \frac{4}{\varepsilon^2} \mathbb{E}\left[\sup_{s \leq S'} \left| \frac{N_s^d}{d} - s \right|^2\right] \leq \frac{16}{\varepsilon^2} \mathbb{E}\left[\left| \frac{N_{S'}^d - S'}{d} \right|^2\right] = \frac{16}{\varepsilon^2} \frac{dS'}{d^2}$$

which converges to 0, where we used Doob’s inequality. Thus, the claim follows. □

B.2. Sufficient conditions for consistency

The following lemma is a fundamental result for consistency of MCMC.

Lemma B.2 (Lemma 2 of [11]). *Let $X^d = \{X_m^d\}_m$ be a sequence of stationary processes on \mathbb{R}^k . If X^d converges in law to $X = \{X_m\}_m$, and if X is a stationary ergodic process, then*

$$\frac{1}{M} \sum_{m=0}^{M-1} f(X_m^d) - \mathbb{E}[f(X_0^d)] = o_{\mathbb{P}}(1) \quad (M, d \rightarrow \infty)$$

for any bounded continuous function $f : \mathbb{R}^k \rightarrow \mathbb{R}$.

Proof. Since $\mathcal{L}(X_0^d) \rightarrow \mathcal{L}(X_0)$, we can substitute $\mathbb{E}[f(X_0^d)]$ by $\mathbb{E}[f(X_0)]$ in the above equation, and hence it is sufficient to show

$$\mathbb{E}\left[\left|\frac{1}{M} \sum_{m=0}^{M-1} f(X_m^d)\right|\right] = o(1) \quad (M, d \rightarrow \infty)$$

for f such that $\mathbb{E}[f(X_0)] = 0$. For such f and any $\varepsilon > 0$, choose $M_0 \in \mathbb{N}$ so that

$$\mathbb{E}\left[\left|\frac{1}{M_0} \sum_{m=0}^{M_0-1} f(X_m)\right|\right] \leq \varepsilon.$$

Then, by stationarity,

$$\begin{aligned} \mathbb{E}\left[\left|\frac{1}{M} \sum_{m=0}^{M-1} f(X_m^d)\right|\right] &\leq \mathbb{E}\left[\left|\frac{1}{M} \sum_{k=0}^{[M/M_0]-1} \sum_{m=0}^{M_0-1} f(X_{M_0k+m}^d)\right|\right] + \mathbb{E}\left[\left|\frac{1}{M} \sum_{m=[M/M_0]M_0}^{M-1} f(X_m^d)\right|\right] \\ &\leq \frac{M_0}{M} \left[\frac{M}{M_0}\right] \mathbb{E}\left[\left|\frac{1}{M_0} \sum_{m=0}^{M_0-1} f(X_m^d)\right|\right] + \|f\|_{\infty} \frac{M - [M/M_0]M_0}{M} \\ &\rightarrow \mathbb{E}\left[\left|\frac{1}{M_0} \sum_{m=0}^{M_0-1} f(X_m)\right|\right] \leq \varepsilon \quad (M, d \rightarrow \infty). \end{aligned}$$

Thus, the claim follows. □

We need a generalization of this lemma. Let $k_1, k_2 \in \mathbb{N}$. Suppose that $\mathbb{R}^{k_1+k_2}$ -valued random variable X_m^d has two parts, $X_m^d = (X_m^{d,1}, X_m^{d,2})$ where $X_m^{d,i}$ is \mathbb{R}^{k_i} valued for each $i = 1, 2$. Corresponding to $X^{d,1}$ and $X^{d,2}$, the invariant probability measure has the following decomposition

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

Furthermore, we assume the following. Let $T_d \rightarrow \infty$.

Assumption B.1. 1. For $Y_t^d = X_{[T_d t]}^{d,1}$, $Y^d \Rightarrow Y$ (in Skorohod’s sense) where Y is stationary and ergodic process with the invariant probability measure P^1 .

2. Random variables $X^d = \{X_m^d\}_m$ converges to $X = \{X_m\}_m = \{(\xi, X_m^2)\}_m$ where $\xi \sim P^1$ and conditioned on ξ , the process $X^2 = \{X_m^2\}_m$ is stationary and ergodic with the invariant probability measure $P^{2|1}(\cdot|\xi)$.

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

Lemma B.3. Let $X^d = \{X_m^d = (X_m^{d,1}, X_m^{d,2})\}_m$ be a sequence of stationary processes on $\mathbb{R}^{k_1+k_2}$. Under the above assumption, for any continuous and bounded function f

$$\frac{1}{M_d} \sum_{m=0}^{M_d-1} f(X_m^d) - P_d(f) = o_{\mathbb{P}}(1)$$

for $M_d \rightarrow \infty$ such that $M_d/T_d \rightarrow \infty$.

Proof. As in the previous lemma, it is sufficient to show

$$\mathbb{E} \left[\left| \frac{1}{T_d M} \sum_{m=0}^{T_d M-1} f(X_m^d) \right| \right] \rightarrow 0 \quad (M, d \rightarrow \infty)$$

for f such that $\int f(x)P^{2|1}(dx_2|x_1)P(dx_1) = 0$. For such f and for $\varepsilon > 0$, choose M_0 so that

$$\mathbb{E} \left[\left| \frac{1}{M_0} \int_0^{M_0} g(Y_t) dt \right| \right] < \varepsilon/2,$$

where $g(x_2) = P^{2|1} f(x_2)$. Then as in the previous lemma,

$$\begin{aligned} & \mathbb{E} \left[\left| \frac{1}{T_d M} \sum_{m=0}^{T_d M-1} g(X_m^{d,1}) \right| \right] \\ &= \mathbb{E} \left[\left| \frac{1}{M} \int_0^M g(Y_t^d) dt \right| \right] \end{aligned}$$

$$\begin{aligned} &\leq \mathbb{E} \left[\left| \frac{1}{M} \sum_{k=0}^{\lfloor M/M_0 \rfloor - 1} \int_{kM_0}^{(k+1)M_0} g(Y_t^d) dt \right| \right] + \mathbb{E} \left[\left| \frac{1}{M} \int_{\lfloor M/M_0 \rfloor M_0}^M g(Y_t^d) dt \right| \right] \\ &\leq \frac{M_0}{M} \left[\frac{M}{M_0} \right] \mathbb{E} \left[\left| \frac{1}{M_0} \int_0^{M_0} g(Y_t^d) dt \right| \right] + \mathbb{E} \left[\left| \frac{1}{M} \int_{\lfloor M/M_0 \rfloor M_0}^M g(Y_t^d) dt \right| \right] \\ &\rightarrow \mathbb{E} \left[\left| \frac{1}{M_0} \int_0^{M_0} g(Y_t) dt \right| \right] \leq \varepsilon/2. \end{aligned}$$

We still need to show

$$\limsup_{d, M \rightarrow \infty} \mathbb{E} \left[\left| \frac{1}{T_d M} \sum_{m=0}^{T_d M - 1} f(X_m^d) - \frac{1}{T_d M} \sum_{m=0}^{T_d M - 1} g(X_m^{d,1}) \right| \right] \leq \varepsilon/2.$$

Set $S_d = T_d M$. By replacing $f(x)$ by $f(x_1, x_2) - g(x_1)$, we can assume $g \equiv 0$. Choose $S_0 \in \mathbb{N}$ so that

$$\mathbb{E} \left[\left| \frac{1}{S_0} \sum_{m=0}^{S_0 - 1} f(X_m) \right| \right] = \mathbb{E} \left[\left| \frac{1}{S_0} \sum_{m=0}^{S_0 - 1} f(\xi, X_m^2) \right| \right] \leq \varepsilon/2.$$

Then, as in the previous lemma, we can show that

$$\begin{aligned} \mathbb{E} \left[\left| \frac{1}{S_d} \sum_{m=0}^{S_d - 1} f(X_m^d) \right| \right] &\leq \frac{S_0}{S_d} \left[\frac{S_d}{S_0} \right] \mathbb{E} \left[\left| \frac{1}{S_0} \sum_{m=0}^{S_0 - 1} f(X_m^d) \right| \right] + \|f\|_\infty \frac{S_d - \lfloor S_d/S_0 \rfloor S_0}{S_d} \\ &\rightarrow \mathbb{E} \left[\left| \frac{1}{S_0} \sum_{m=0}^{S_0 - 1} f(X_m) \right| \right] \leq \varepsilon/2. \end{aligned}$$

Thus, we can conclude that

$$\begin{aligned} &\limsup_{d, M \rightarrow \infty} \mathbb{E} \left[\left| \frac{1}{T_d M} \sum_{m=0}^{T_d M - 1} f(X_m^d) \right| \right] \\ &\leq \limsup_{d, M \rightarrow \infty} \left\{ \mathbb{E} \left[\left| \frac{1}{T_d M} \sum_{m=0}^{T_d M - 1} g(X_m^{d,1}) \right| \right] + \mathbb{E} \left[\left| \frac{1}{T_d M} \sum_{m=0}^{T_d M - 1} (f(X_m^d) - g(X_m^{d,1})) \right| \right] \right\} \\ &\leq \varepsilon. \end{aligned}$$

Hence, the proof is completed. □

Acknowledgement

The author wishes to thank to Andreas Eberle, Ajay Jasra, Gareth O. Roberts and Masayuki Uchida for fruitful discussions. A part of this work was done when the author was visiting the

Institute for Applied Mathematics, Bonn University. The author thanks the Institute for Applied Mathematics, Bonn University for its hospitality. He also thanks to the former Editor Professor Eric Moulines, an anonymous Associate Editor and anonymous referees for their extremely careful readings of this manuscript and their many insightful comments which lead to numerous improvements. This work was supported in part by JSPS KAKENHI Grant Numbers 24740062, 16K00046 and JST CREST Grant Number JPMJCR14D7.

References

- [1] Beskos, A., Roberts, G. and Stuart, A. (2009). Optimal scalings for local Metropolis–Hastings chains on nonproduct targets in high dimensions. *Ann. Appl. Probab.* **19** 863–898. [MR2537193](#)
- [2] Chen, L.H.Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein’s Method. Probability and Its Applications*. Springer, Berlin.
- [3] Cotter, S.L., Roberts, G.O., Stuart, A.M. and White, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28** 424–446. [MR3135540](#)
- [4] Eberle, A. (2014). Error bounds for Metropolis–Hastings algorithms applied to perturbations of Gaussian measures in high dimensions. *Ann. Appl. Probab.* **24** 337–377. [MR3161650](#)
- [5] Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence*. New York: Wiley. [MR0838085](#)
- [6] Geyer, C.J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* **7** 473–483.
- [7] Goto, F. (2017). An extension and practical performances of mpcn algorithm. Master’s thesis, Graduate School of Engineering Science, Osaka University.
- [8] Hairer, M., Stuart, A.M. and Vollmer, S.J. (2014). Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.* **24** 2455–2490. [MR3262508](#)
- [9] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Berlin: Springer. [MR1943877](#)
- [10] Kamatani, K. (2014). Rate optimality of Random walk Metropolis algorithm in high-dimension with heavy-tailed target distribution. Preprint. Available at [arXiv:1406.5392](#).
- [11] Kamatani, K. (2014). Local consistency of Markov chain Monte Carlo methods. *Ann. Inst. Statist. Math.* **66** 63–74. [MR3147545](#)
- [12] Kamatani, K. (2017). Ergodicity of Markov chain Monte Carlo with reversible proposal. *J. Appl. Probab.* **54** 638–654. [MR3668488](#)
- [13] Kamatani, K., Nogita, A. and Uchida, M. (2016). Hybrid multi-step estimation of the volatility for stochastic regression models. *Bull. Inform. Cybernet.* **48** 19–35. [MR3644196](#)
- [14] Kamatani, K. and Uchida, M. (2015). Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Stat. Inference Stoch. Process.* **18** 177–204. [MR3348584](#)
- [15] Karatzas, I. and Shreve, S.E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. New York: Springer. [MR1121940](#)
- [16] Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge: Cambridge Univ. Press. [MR2038227](#)
- [17] Neal, R.M. (1999). Regression and classification using Gaussian process priors. In *Bayesian Statistics, 6 (Alcoceber, 1998)* 475–501. New York: Oxford Univ. Press. [MR1723510](#)
- [18] Nourdin, I. and Peccati, G. (2009). Stein’s method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. [MR2520122](#)

- [19] Nourdin, I. and Peccati, G. (2012). *Normal Approximations with Malliavin Calculus, from Stein's Method to Universality*. *Cambridge Tracts in Mathematics* **192**. Cambridge: Cambridge Univ. Press. [MR2962301](#)
- [20] Nualart, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. Berlin: Springer. [MR2200233](#)
- [21] Pillai, N.S., Stuart, A.M. and Thiery, A.H. (2014). Optimal proposal design for random walk type Metropolis Algorithms with Gaussian random field priors. Preprint. Available at [arXiv:1108.1494v2](#).
- [22] Plummer, M., Best, N., Cowles, K. and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News* **6** 7–11.
- [23] Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer. [MR2080278](#)
- [24] Roberts, G.O., Gelman, A. and Gilks, W.R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. [MR1428751](#)
- [25] Roberts, G.O. and Rosenthal, J.S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. [MR1625691](#)
- [26] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. [MR1739520](#)
- [27] Shigekawa, I. (1980). Derivatives of Wiener functionals and absolute continuity of induced measures. *J. Math. Kyoto Univ.* **20** 263–289. [MR0582167](#)
- [28] Stroock, D.W. and Varadhan, S.R.S. (1979). *Multidimensional Diffusion Processes*. *Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen Mit Besonderer Berücksichtigung der Anwendungsgebiete*. Berlin: Springer.
- [29] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. [MR1329166](#)

Received January 2015 and revised March 2017