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 and CREST, JST

Dated: March 12, 2017

Abstract

The purpose of this paper is to introduce a new Markov chain Monte Carlo method and to expose its effectiveness by simulation and high-dimensional asymptotic theory. The key fact is that our algorithm has a reversible proposal transition kernel, which is designed to have a heavy-tailed invariant probability distribution. A high-dimensional asymptotic theory is studied for a heavy-tailed target probability distribution class. 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: Markov chain; Consistency; Monte Carlo; Stein's method; Malliavin calculus

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 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 the heavy-tailed target probability distribution.

In our previous work, in Kamatani [2014b], we studied some asymptotic properties of the random-walk Metropolis (RWM) algorithm for heavy-tailed target probability distribution. To perform the RWM algorithm, we have to choose a propositional distribution. This choice strongly affects performance. We showed that the most usual choice, the RWM Gaussian algorithm has a slow convergence rate for the case of the heavy tail, even if the choice reaches the optimal rate among the proposed distributions. This rather disappointing fact illustrates that the RWM algorithm can not be so good. 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 Neal [1999]. 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 Cotter et al. [2013] and its theoretical effect was provided in Beskos et al. [2009], Pillai et al. [2014], Eberle [2014] and Hairer et al. [2014]. However, our simulation shows that it works only for a specific light distribution and does not work well otherwise, especially for the heavy-tailed target probability distribution. In Theorem 3.1, we will prove it 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 and is very robust. Let us describe our

*Supported in part by Grant-in-Aid for Young Scientists (B) 24740062.

new algorithm, the mixed pre-conditioned Crank-Nicolson algorithm (MpCN). Let $P(\mathrm{d}x) = p(x)\mathrm{d}x$ be the target probability distribution on $\mathbb{R}^d$. Fix $\rho \in (0,1)$. Set initial value $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and let $\|x\| = (\sum_{i=1}^d x_i^2)^{1/2}$. The algorithm goes as follows:

- Generate $r \sim \mathrm{Gamma}(d/2, \|x\|^2/2)$.
- 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, $\mathrm{Gamma}(\nu, \alpha)$ is the Gamma distribution with probability density function $g(x;\nu,\alpha) = x^{\nu-1}\alpha^{\nu}\exp(-\alpha x)/\Gamma(\nu)$.

The key fact is that the transition kernel of the proposed algorithm has the heavy-tailed invariant probability distribution. Thus, it is not surprising that the new method works better than the pCN algorithm for the heavy-tailed target probability distribution. 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 is robust, which is one of the most important property for MCMC.

We study its theoretical properties via high-dimensional asymptotic theory. The high-dimensional asymptotic theory for MCMC was first appeared in Roberts et al. [1997] and further developed in Roberts and Rosenthal [1998]. See Cotter et al. [2013] for recent results. We use this framework together with the study of consistency of MCMC by Kamatani [2014a].

The main technical tools are Malliavin calculus and Stein's techniques. The reader is referred to Nualart [2006] for the former and Chen et al. [2011] for the latter and see Nourdin and Poly [2013] 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 Bayesian computation.

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 heavy-tailed target probability distribution. On the other hand, the MpCN algorithm attains a better rate than the RWM algorithm. Proofs are relegated to Section 4. In the appendix, Section A includes a short introduction to Malliavin calculus and Stein's techniques. Section B provides some properties for consistency of MCMC.

Finally, we note that our new algorithm was already implemented for the Bayesian type estimation for ergodic diffusion process in Kamatani and Uchida [2015] and diffusion process with high-frequency data Kamatani et al. [2016] (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 Kamatani and Uchida [2015]. In our current study, we only describe the usefulness of our algorithm for a class of heavy-tailed target probability distribution. However, this heavy tail assumption is just an example of target probability distribution that is difficult to approximate by MCMC. Our method is robust, and we believe that the method is also useful for not heavy-tailed complicated target probability distribution as illustrated in Kamatani and Uchida [2015] and Kamatani et al. [2016].

1.1 Notation

Several norms are considered in this paper.

- For $x = (x_1, \ldots, x_d),\ y = (y_1, \ldots, y_d) \in \mathbb{R}^d$, write $\|x\| = \left(\sum_{i=1}^d x_i^2\right)^{1/2}$ and $\langle x, y\rangle = \sum_{i=1}^d x_i y_i$.
- For a function $f : E \to \mathbb{R}$, write $\|f\|_\infty = \sup_{x\in E} |f(x)|$.

- If ν is a signed measure on $(E, \mathcal{E})$, write $\|\nu\|_{\mathrm{TV}} = \sup_{A \in \mathcal{E}} |\nu(A)|$. The integral with respect to ν is denoted by $\nu(f) = \nu f = \int_E f(x)\nu(\mathrm{d}x)$.
- $[x]$ is the integer part of $x > 0$.
- Write $N_d(\mu, \Sigma)$ for the d-dimensional normal distribution with mean $\mu \in \mathbb{R}^d$ and variance covariance matrix Σ, and $\phi_d(x; \mu, \Sigma)$ be its probability density function. When $d = 1$, write $N(\mu, \sigma^2)$ and $\phi(x; \mu, \sigma^2)$ with respectively.
- Write I_d for the $d \times d$-identity matrix.
- Write $\mathcal{L}(X)$ for the law of random variable X. Write $X_n \Rightarrow X$ if the law of X_n converges weakly to that of X. Write $X_n = O_{\mathbb{P}}(1)$ when the sequence $\mathcal{L}(X_n)$ $(n = 1, 2, \ldots,)$ is tight, 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 $P(\mathrm{d}x)$ be a probability measure on $(E, \mathcal{E})$ with a probability density function $p(x)$ with respect to a σ-finite measure $\nu(\mathrm{d}x)$. The Metropolis-Hastings algorithm generates a Markov chain $\{X_m\}_m$ with transition kernel $K(x, \mathrm{d}y)$ on $(E, \mathcal{E})$ defined by the following: Set $X_0 \in E$ and for $m \geq 1$,

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

where $R(x, \mathrm{d}y)$ is called the proposal transition kernel, and $\alpha(x, y)$ is called the acceptance ratio which is defined by

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

where $\mathcal{E}^{\otimes 2}$-measurable function $r(x, y)$ is the probability density function of $R(x, \mathrm{d}y)$. This acceptance probability is designed to satisfy

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

The Markov chain is called reversible with respect to $P(\mathrm{d}x)$ if

$$P(\mathrm{d}x)K(x, \mathrm{d}y) = P(\mathrm{d}y)K(y, \mathrm{d}x).$$

If the acceptance ratio satisfies (2.1), then the Markov chain has reversibility. See monograph Robert and Casella [2004] or review Tierney [1994] for further details.

When $E = \mathbb{R}^d$, the most popular choice is $R(x, \mathrm{d}y) = N_d(x, \Sigma)$ where Σ is a positive definite matrix. However, this popular choice reveals the limitation in high-dimension as described in Roberts et al. [1997]. Another approach was proposed by Neal [1999] which sometimes works better. Let P_d be a probability measure on $\mathbb{R}^d$ with density $p_d(x)$. 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 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 transition kernel is $R_d(x,\mathrm{d}y)=N_d(\sqrt{\rho}x,(1-\rho)I_d)$. The probability transition kernel $R_d(x,\mathrm{d}y)$ is the conditional distribution $Y|X$ of the following joint distribution:

$$(X,Y)\sim N_{2d}\left(0,\begin{pmatrix} I_d & \sqrt{1-\rho}I_d\\ \sqrt{1-\rho}I_d & I_d\end{pmatrix}\right).$$

By this fact, the invariant distribution of R_d is $N_d(0,I_d)$. In particular, if $P_d=N_d(0,I_d)$ and $X_0^d\sim P_d$, each X_m^d is always accepted and it becomes a d-dimensional AR(1) process.

The pCN algorithm works well when the target distribution P_d is approximately Gaussian distribution. However, we will see that the algorithm becomes even worse when the target distribution has a heavy tail. This property is due to the invariant distribution of R_d is a Gaussian distribution.

2.2 The MpCN algorithm

In this paper, we propose the following algorithm that generate a Markov chain $X^d=\left\{X_m^d\right\}_{m\in\mathbb{N}_0}$: 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

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

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

Formally, the proposal transition kernel $R_d(x,\mathrm{d}y)$ of the MpCN algorithm is the conditional distribution $Y|X$ of the following joint distribution:

$$(X,Y)|\sigma^2\sim N_{2d}\left(0,\sigma^2\begin{pmatrix} I_d & \sqrt{1-\rho}I_d\\ \sqrt{1-\rho}I_d & I_d\end{pmatrix}\right),\sigma^2\sim\overline{Q}$$

when $\overline{Q}(\mathrm{d}x)=1_{(0,\infty)}(x)x^{-1}\mathrm{d}x$ is a 'prior distribution' for σ^2. It is not difficult to check this since

$$\begin{aligned}\sigma^{-2}|X &\sim \text{Gamma}(d/2,\|X\|^2/2),\\ Y|X,\sigma^2 &\sim N_d(\sqrt{\rho}X,(1-\rho)\sigma^2 I_d).\end{aligned}$$

By this structure, by integrating out σ^2 from the joint distribution of (X,σ^2), the transition kernel is reversible with respect to

$$\overline{P}_d(\mathrm{d}x)\propto\int_{\sigma^2\in(0,\infty)}\phi_d(x,0,\sigma^2 I_d)\overline{Q}(\mathrm{d}\sigma^2)\mathrm{d}x\propto\|x\|^{-d}\mathrm{d}x. \tag{2.3}$$

Since $\overline{P}_d$ has a heavier tail than those of Gaussian distributions, we expect that this new method works well even for heavy-tailed target 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 will compare two RWM algorithms and the pCN and MpCN algorithms with $M = 10^8$ iterations (no burn-in) for each. The algorithms we consider are

1. The RWM algorithm with Gaussian proposal distribution. More precisely, the update x^* from the current value x is generated by $x^* = x + \sigma_d \epsilon$ 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, $x^* = x + \sigma_d \epsilon$ 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.

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.

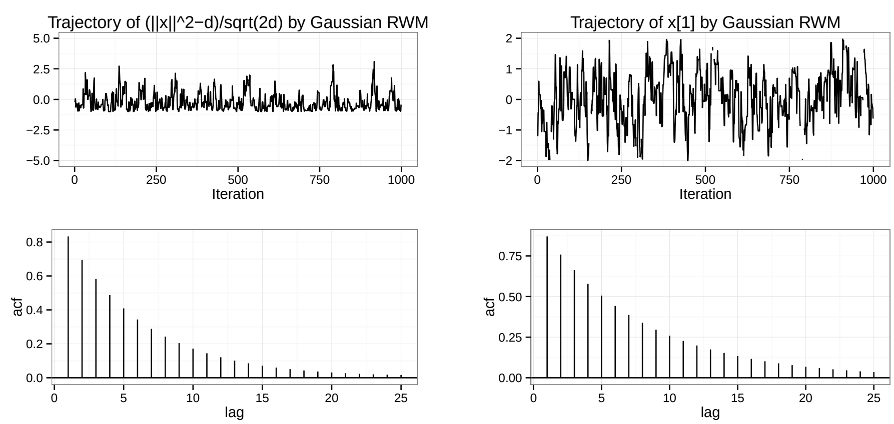

Figure 1: The RWM algorithm with Gaussian proposal distribution for $P_d = N_d(0, I_d)$ for $d = 2$.

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 $\{X^d_{m,1}\}_m$ where $X^d_m = (X^d_{m,1}, \ldots, X^d_{m,d})$ (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 $P_d(\xi\mathbf{1} - \mathrm{d}x)$ where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^d$ and

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

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 $\left\{(2d)^{-1/2}(\|X^d_m - \xi\mathbf{1}\|^2 - d)\right\}_m$ for the standard normal distribution, and plot that of $\left\{\|X^d_m - \xi\mathbf{1}\|^2/d\right\}_m$ 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 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 distribution. We will show the results with and without this estimation.

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 optimal convergence rate for the RWM algorithm is d, and the Gaussian proposal distribution attains this rate (Theorem 3.1 of Kamatani [2014b]). On the other hand, the pCN and MpCN algorithms attain consistency, and so these algorithms are better than the optimal 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.

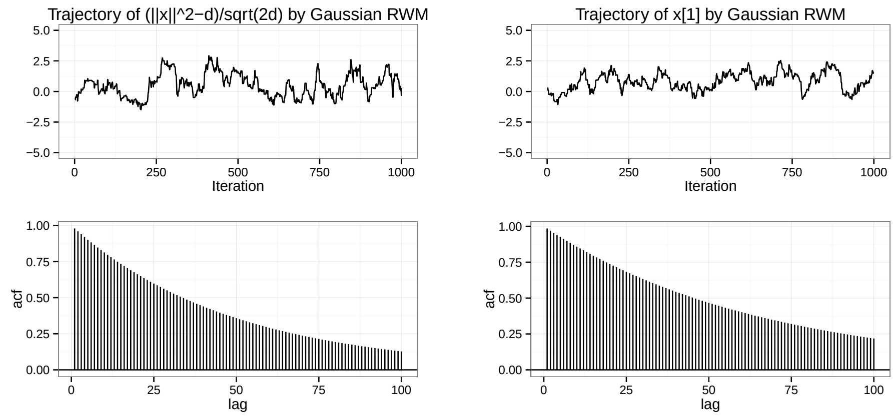

Figure 2: The RWM algorithm with Gaussian proposal distribution for $P_d = N_d(0, I_d)$ for $d = 20$.

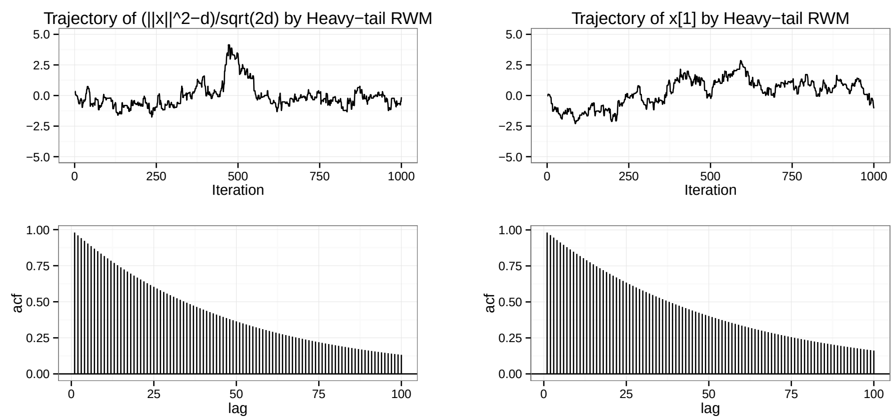

Figure 3: The RWM algorithm with t-distribution as the proposal distribution for $P_d = N_d(0, I_d)$ for $d = 20$.

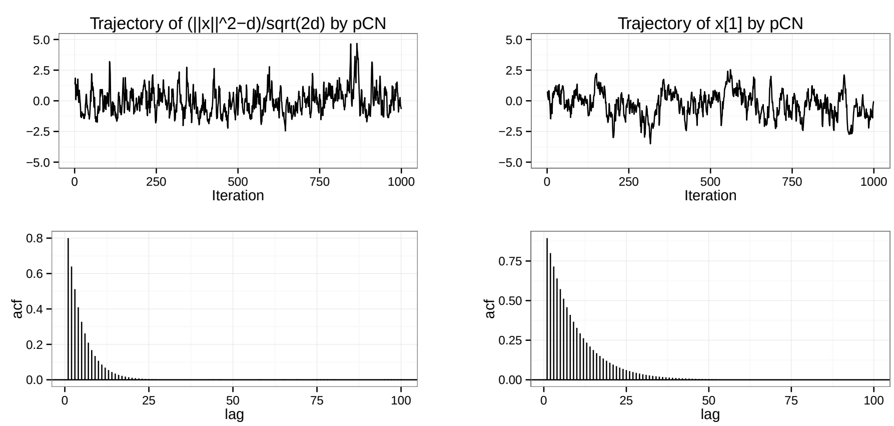

Figure 4: The pCN algorithm for $P_d = N_d(0, I_d)$ for $d = 20$.

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

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

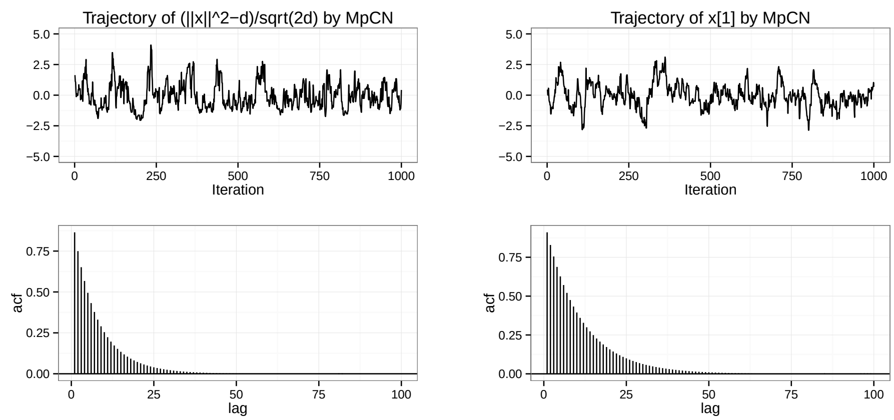

Figure 5: The MpCN algorithm for $P_d = N_d(0, I_d)$ for $d = 20$.

We also observed effective sample size (ESS; Geyer [1992]. See also 12.3.5 of Robert and Casella [2004]) by using coda package (Plummer et al. [2006]) in R. The results are calculated from 5000 samples after 5000 burn-in samples for 50 parallel runs and calculated the average over d coordinates. The value of 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.

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 location parameter $\mu = 0$ and the scale parameter $\sigma = 5$ for $d = 20$. Recall that the probability distribution function is given by

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

For this case, the optimal convergence rate for the RWM algorithm is d^2, and the Gaussian proposal distribution attains this rate (Theorem 3.2 of Kamatani [2014b]). The pCN algorithm is much worse than the rate, and the MpCN algorithm attains much better rate d (Theorems 3.1 and 3.3). In the simulation, the MpCN algorithm (Figure 9) is much better than other algorithms (Figures 6-8) which correspond to the theoretical result.

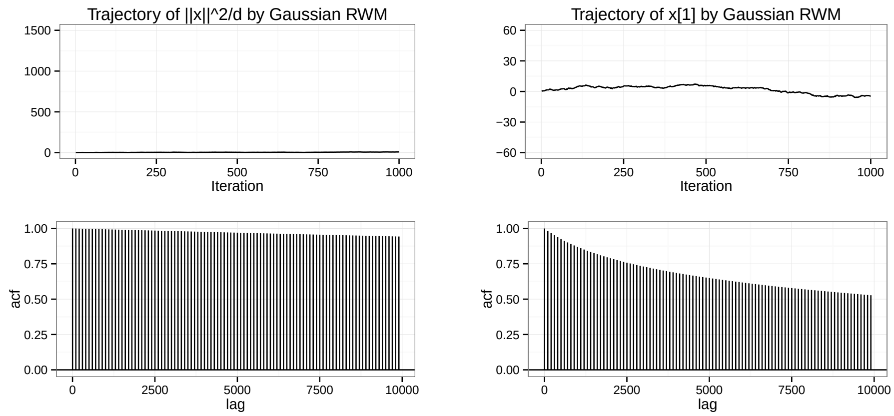

Figure 6: The RWM algorithm with Gaussian proposal distribution when t-distribution is the target distribution.

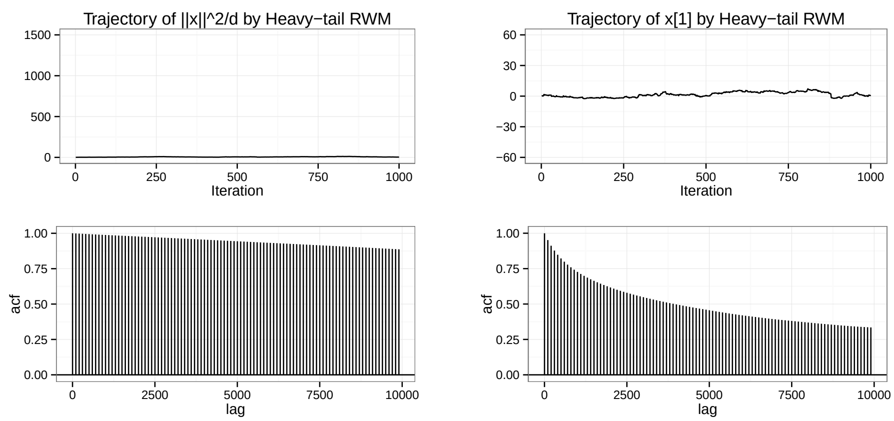

Figure 7: The RWM algorithm with t-distribution as the proposal distribution and the target distribution is also the t-distribution.

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

Table 2: Effective sample size for t-distribution

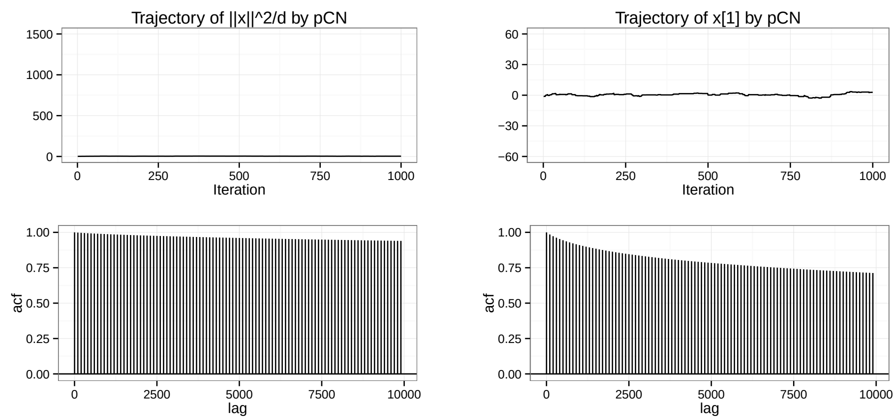

Figure 8: The pCN algorithm when t-distribution is the target probability distribution.

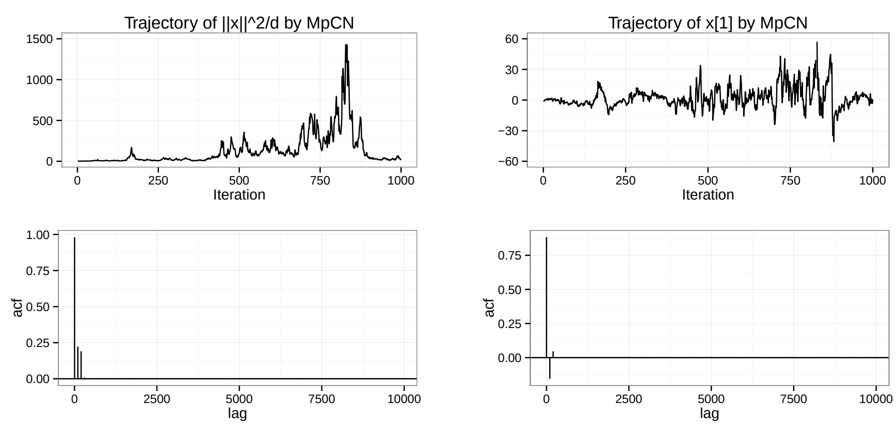

Figure 9: The MpCN algorithm when t-distribution is the target probability distribution.

The estimated value of ESS also reveals the limitation of pCN, and efficiency of 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 distribution is not spherically symmetric. Let P_{20} be a probability measure in $\mathbb{R}^{20}$ with the probability density function

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

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 (Figure 13) is much better than other algorithms (Figures 10-12).

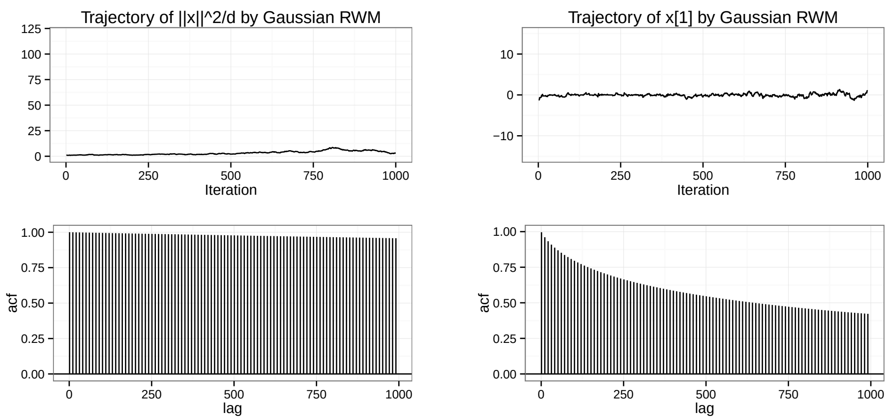

Figure 10: The RWM algorithm with Gaussian proposal distribution when the perturbed t-distribution is the target probability distribution.

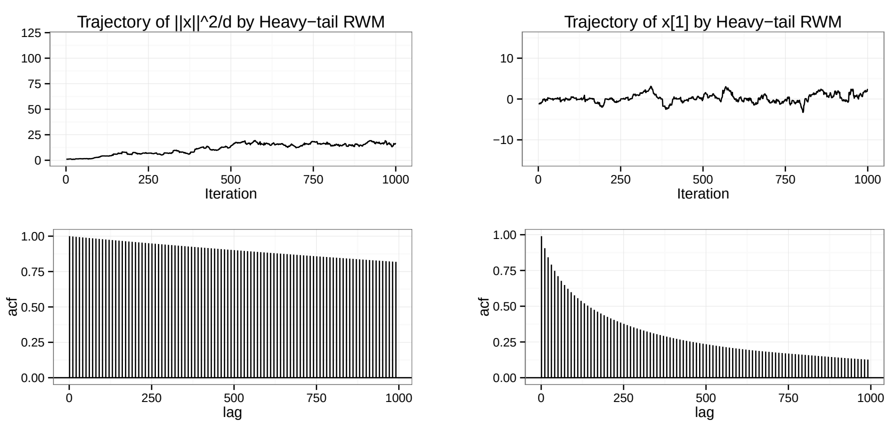

Figure 11: The RWM algorithm with t-distribution as the proposal distribution and the target probability distribution is the perturbed t-distribution.

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

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

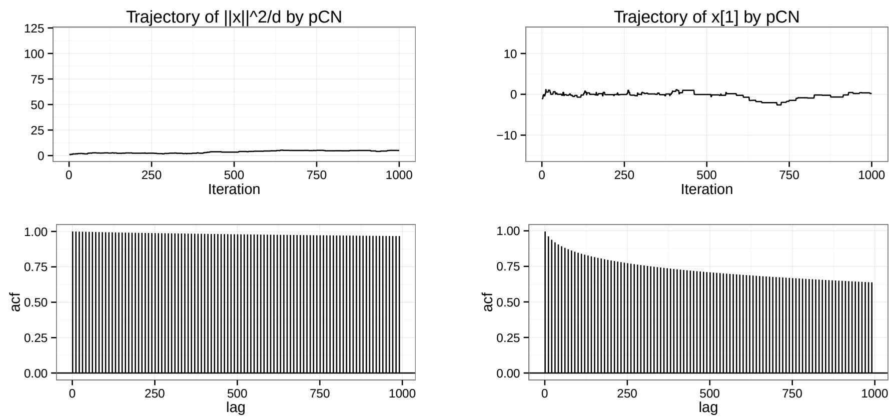

Figure 12: The pCN algorithm when the perturbed t-distribution is the target probability distribution.

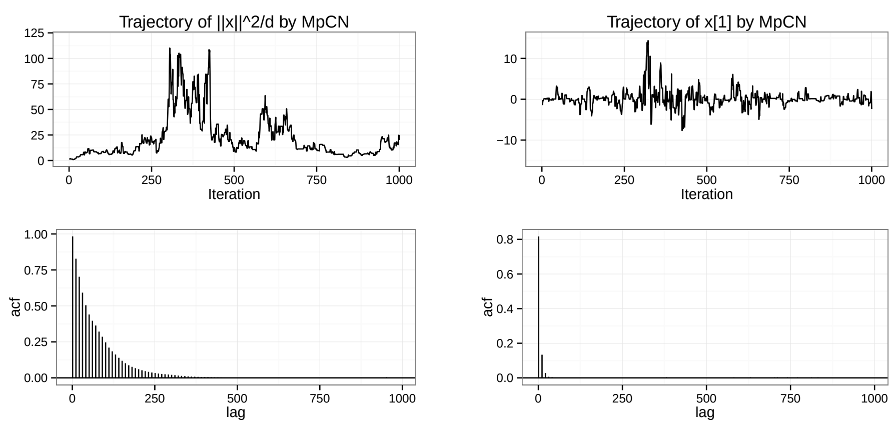

Figure 13: The MpCN algorithm when the perturbed t-distribution is the target probability distribution.

Again, we calculated the estimated ESS. 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 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 Figures 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 distribution in high-dimension, when the high-probability region is far from the origin, it is important to shift the target distribution in advance. For example, first, calculate rough estimate $\hat{\xi}$ of the center of the target distribution $P_d(\mathrm{d}x)$, and then run the MCMC algorithm for $P_d(-\hat{\xi} + \mathrm{d}x)$. Some tempering strategy might be useful for the rough estimate of the center of the target distribution as in Kamatani and Uchida [2015].

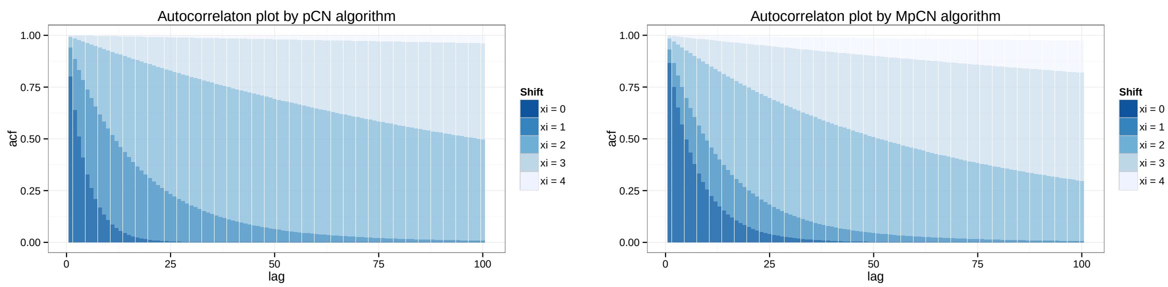

Figure 14: Autocorrelation plots for the pCN and MpCN algorithms for shifted normal distributions.

Next figure (Figure 15) is a result of the pCN and MpCN algorithm with a simple estimation of the center of the target distribution. We run $M = 10^3$ iteration of the pCN or MpCN 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 $P_d(-\hat{\xi} + \mathrm{d}x)$. The effect of the shift is considerably weakened.

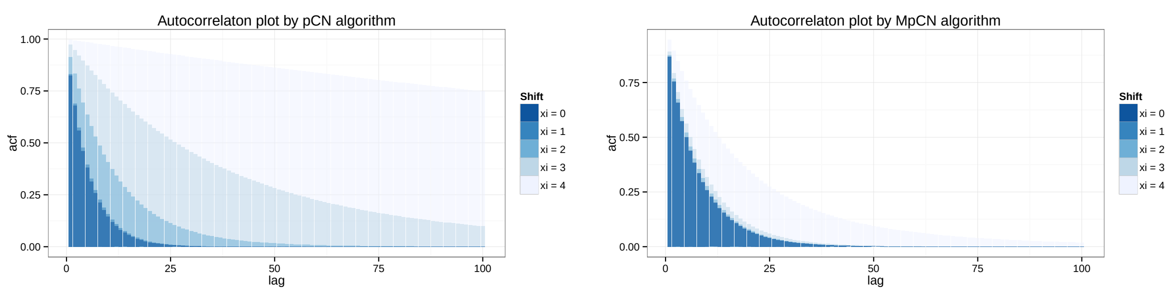

Figure 15: Autocorrelation plots for the pCN and MpCN algorithms for shifted normal distributions with an initial estimate of the center of the target distribution.

We consider the t-distribution with $\nu = 2$ and $\sigma = 5$ 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.

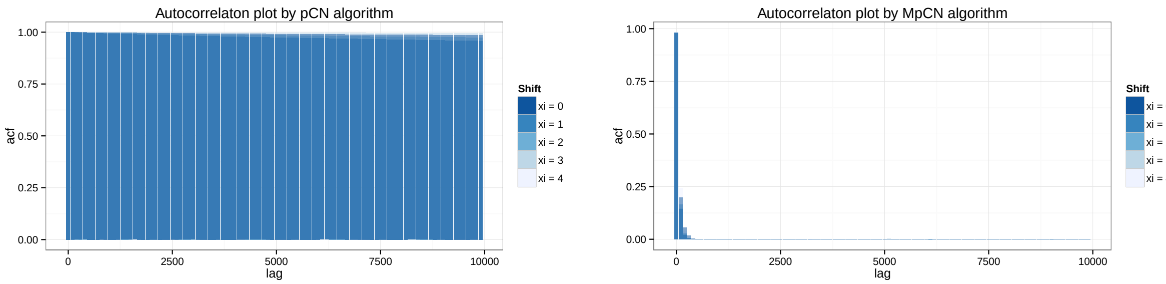

Figure 16: Autocorrelation plots for the pCN and MpCN algorithms for shifted t-distributions.

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 distribution (2.4) by $M = 10^3$ iteration. Thus for heavy-tailed target distribution, the effect of shift perturbation is small, and the gain of the estimation of the center is also small.

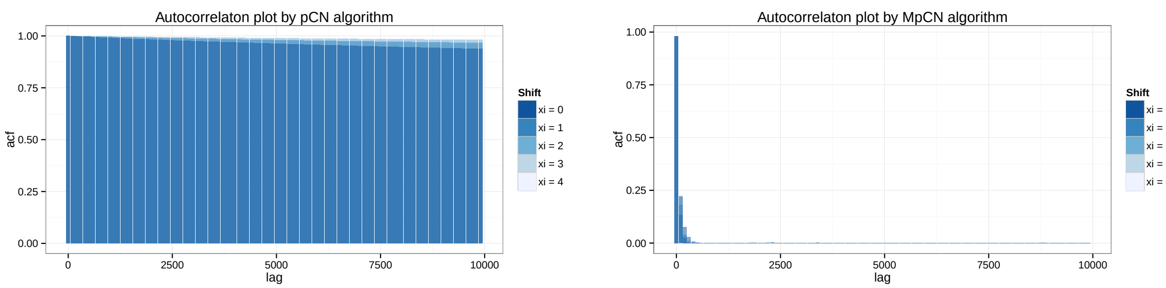

Figure 17: Autocorrelation plots for the pCN and MpCN algorithms for shifted t-distributions with an initial estimate of the center of the probability measure.

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 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 are essential. However, this can be weakened as explained in Lemma 4 of Kamatani [2014a].

3.1 Consistency

In this section, we review the consistency of MCMC studied in Kamatani [2014a]. For each $d \in \mathbb{N}$, suppose that Markov chain Monte Carlo method $\mathscr{M}^d$ generates a Markov chain $\left\{X_m^d; m \in \mathbb{N}_0\right\}$ with the invariant probability distribution Π_d. A family $\mathscr{M}^d$ ($d \in \mathbb{N}$) is called consistent if

$$\frac{1}{M}\sum_{m=0}^{M-1} f(X_m^d) - \Pi_d(f) = o_{\mathbb{P}}(1) \tag{3.1}$$

for any $M, d \to \infty$ for any bounded continuous function f.

If the family $\mathscr{M}^d$ does not depend on d, this is just a weak law of large numbers. The consistency says that the integral $\Pi_d(f)$ 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 Kamatani [2014a]) when d is the sample size of the data.

In the current context, the state space for $X^d = \left\{X^d_m; m \in \mathbb{N}_0\right\}$ $(d \in \mathbb{N})$ changes as $d \to \infty$ that is inconvenient for further analysis. As in Roberts et al. [1997] and Kamatani [2014b], to overcome the difficulty, we set a projection $\pi_k = \pi_{d,k}$ to a finite subset by

$$\pi_k(x) = (x_i)_{i=1,\ldots,k}\ (x = (x_i)_{i=1,\ldots,d}).$$

Other possibility to overcome this difficulty is to embed $\mathbb{R}^d$ into $\mathbb{R}^{\mathbb{N}}$ under suitable metric as in Beskos et al. [2009]. 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 1 (Consistency)**.** *A family $\mathscr{M}^d$ $(d \in \mathbb{N})$ is called consistent if for any $k \in \mathbb{N}$, $M_d \to \infty$ and for any bounded continuous function $f : \mathbb{R}^k \to \mathbb{R}$,*

$$\frac{1}{M_d}\sum_{m=0}^{M_d-1} f \circ \pi_k(X^d_m) - P_d(f \circ \pi_k) = o_{\mathbb{P}}(1)\ (d \to \infty). \tag{3.2}$$

We will see that the pCN and MpCN algorithms are consistent for light-tailed target distribution. However, when P_d is a heavy-tailed distribution, these methods do not have consistency, but weak consistency defined by the following.

Definition 2 (Weak Consistency)**.** *A family $\mathscr{M}^d$ $(d \in \mathbb{N})$ is called weakly consistent if with rate T_d if (3.2) is satisfied for any $M_d \to \infty$ such that $M_d/T_d \to \infty$. We will call the rate T_d, the convergence rate. If $T_d/d^k \to 0$ 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 $\mathscr{M}^d$ is consistent, then the convergence rate is $T_d = 1$.

Example 3.1. *Let $\mathscr{M}^d$ be the random-walk Metropolis algorithm with target 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, $X^d_0 \sim P_d$ for the Markov chain $\{X^d_m\}_m$ generated by $\mathscr{M}^d$. Then as in Theorem 1.1 of Roberts et al. [1997], for any $k \in \mathbb{N}$, $Y^d_t := \pi_k(X^d_{[dt]})$ converges weakly to a k-dimensional Ornstein-Uhlenbeck process. Then $\mathscr{M}^d$ is weakly consistent and the rate is d by Lemma B.3.*

When the target distribution is heavy-tailed, the performance of MCMC algorithms is quite different from that for the light-tailed case. In Kamatani [2014b], we showed that the optimal 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 a class of target distribution considered in this paper. We want to study heavy-tailed target distribution, but the analysis becomes much more complicated than the light-tailed 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 heavy-tailed target distributions such as the t-distribution and stable distribution. See for example Andrews and Mallows [1974] for one-dimensional case.

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

$$P_d = \mathcal{L}(X^d_0),\ Q_d = \mathcal{L}(\|X^d_0\|^2/d) \tag{3.3}$$

where $X_0^d|\sigma^2 \sim N_d(0, \sigma^2 I_d)$ and $\sigma^2 \sim Q$. Note that $Q_d \to Q$ as $d \to \infty$ 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 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*

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

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 Kotz and Nadarajah [2004].

Example 3.3 (Stable distribution)**.** *If P_d is the rotationally symmetric α-stable distribution with characteristic function $\int \exp(-\mathrm{i}\langle t, x\rangle)P_d(\mathrm{d}x) = \exp(-\|t/2\|^{-\alpha})$, then Q is $\alpha/2$-stable distribution on the half line with Laplace transform $\int \exp(-ty)Q(\mathrm{d}y) = \exp(-|t|^{\alpha/2})$. Although there is no closed form of probability density function $q(x)$, all derivatives of $q(x)$ are continuous and vanishes at 0 and ∞. See Section 14 of Sato [1999].*

For this class of target 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

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

We write $\tilde{Q}$ and $\tilde{Q}_d$ for probability measure with densities $\tilde{q}$ and $\tilde{q}_d$. Note that if $\sigma^2 \sim Q$ and $\sigma_d^2 \sim Q^d$, then $(\log\sigma^2)/2 \sim \tilde{Q}$ and $(\log\sigma_d^2)/2 \sim \tilde{Q}_d$. In particular, $\tilde{Q}_d \to \tilde{Q}$.

3.3 Main results

We assume **stationarity** of the process. If the target probability distribution P_d is different from $N_d(0, I_d)$, then any polynomial number of iterations is not sufficient for the pCN algorithm to have a good approximation of the integral we want to calculate.

Theorem 3.1. *The pCN algorithm does not have a polynomial rate of convergence if Assumption 1 is satisfied.*

When P_d is a heavy-tailed distribution, we already know that the pCN algorithm does not work well by Theorem 3.1. However, the MpCN algorithm still has a good convergence property. Recall that the optimal convergence rate for the RWM algorithm is d^2 as studied in Kamatani [2014b]. Let $\eta = \eta(\rho) = \sqrt{1-\rho}/2$. See Section 4 for the proofs.

Proposition 3.2. *Let Q satisfy Assumption 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*

$$\mathrm{d}Y_t = a(Y_t)\mathrm{d}t + \sqrt{b(Y_t)}\mathrm{d}W_t; Y_0 \sim \tilde{Q} \tag{3.5}$$

where

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

Theorem 3.3. *Let Q satisfy Assumption 1. Then the MpCN algorithm has the convergence rate d.*

3.4 Discussion

- Proposal transition kernel $R_d(x, \mathrm{d}y)$ used in MpCN has the invariant distribution $\overline{P}_d$ defined in (2.3), and so this is a special case of MCMC that uses reversible proposal transition kernel. The relation to the target probability distribution P_d and $\overline{P}_d$ is quite important. If $\overline{P}_d$ has a heavier-tail than that of P_d, then MCMC behaves relatively well. On the other hand, if $\overline{P}_d$ has a lighter-tail, it becomes quite poor. The RWM algorithm has $\overline{P}_d =$ Uniform distribution. This is a robust choice, but it loses efficiency to pay the price as described in Kamatani [2014b]. On the other hand, the pCN algorithm, which has $\overline{P}_d = N_d(0, I_d)$, 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 $\overline{Q}$ be any σ-finite measure on $(0, \infty)$ and set $\overline{P}_d$ as in (2.3) with density $\overline{p}_d$. For $m \geq 1$, set

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

  where $\alpha_d(x, y) = \min\{1, p_d(y)\overline{p}_d(x)/p_d(x)\overline{p}_d(y)\}$ assuming that $\int \phi_d(x; 0, zI_d)\overline{Q}(\mathrm{d}z) < \infty$ for any $x \in \mathbb{R}^d$. For example, in Kamatani and Uchida [2015], $\overline{Q}(\mathrm{d}z) \propto z^{-\nu/2-1}e^{-\nu/(2z)}$. See also Goto [2017].

- As studied in Theorem 3.1, the pCN algorithm does not work well if $Q(\{1\}) < 1$. On the other hand, if $Q(\{1\}) = 1$, then P_d becomes the standard normal distribution and the pCN algorithm produces a rejection-free autoregressive process. By this fact, the pCN algorithm has consistency for this case.

- When $P_d = N_d(0, \sigma^2 I_d)$ for $\sigma^2 > 0$, the MpCN algorithm has consistency. 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.

- 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.

- Ergodicity is also studied for this algorithm. The MpCN algorithm can be geometrically ergodic even for heavy-tailed target distributions. See Kamatani [2017] and Goto [2017].

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 $x \in \mathbb{R}^d \backslash \{0\}$ and for independent random variables $W_1^d, \tilde{W}_1^d \sim N_d(0, I_d)$, let

$$F_d(x) = \frac{d^{1/2}}{2\eta}\left\{\log\left\|\sqrt{\rho}x + \sqrt{1-\rho}\|x\|\frac{W_1^d}{\|\tilde{W}_1^d\|}\right\|^2 - \log\|x\|^2\right\}.$$

This random variables essentially determines the behaviors 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 main proof 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 $x \in \mathbb{R}^d \backslash \{0\}$ 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

$$F_d(x) = \frac{d^{1/2}}{2\eta} \log \left\| \sqrt{\rho} \frac{x}{\|x\|} + \sqrt{1-\rho} \frac{W_1^d}{\|\tilde{W}_1^d\|} \right\|^2 = \frac{d^{1/2}}{2\eta} \log \left(\rho + 2\sqrt{\rho(1-\rho)} \frac{\|W_1^d\|}{\|\tilde{W}_1^d\|} U_1^d(x) + (1-\rho) \frac{\|W_1^d\|^2}{\|\tilde{W}_1^d\|^2} \right),$$

where

$$U_1^d(x) = \left\langle \frac{x}{\|x\|}, \frac{W_1^d}{\|W_1^d\|} \right\rangle.$$

In order to prove that $\mathcal{L}(F_d(x))$ does not depend on x, we show that $\mathcal{L}(U_1^d(x))$ has the property. But it is obvious since $\mathcal{L}(U_1^d(x)) = \mathcal{L}(U_1^d(\alpha V x))$ 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

$$F_d := F_d \left(\frac{\tilde{W}_1^d}{\|\tilde{W}_1^d\|} \right) = \frac{d^{1/2}}{2\eta} \left(\log \left\| \sqrt{\rho} \tilde{W}_1^d + \sqrt{1-\rho} W_1^d \right\|^2 - \log \|\tilde{W}_1^d\|^2 \right),$$

where we used the fact that the random variables W_1^d and $\|\tilde{W}_1^d\|$ in $F_d(x)$ are independent from $\tilde{W}_1^d / \|\tilde{W}_1^d\|$. Recall that $(\tilde{W}_1^d, \sqrt{\rho}\tilde{W}_1^d + \sqrt{1-\rho} W_1^d)$ is an exchangeable pair, that is,

$$\mathcal{L}(\tilde{W}_1^d, \sqrt{\rho}\tilde{W}_1^d + \sqrt{1-\rho} W_1^d) = \mathcal{L}(\sqrt{\rho}\tilde{W}_1^d + \sqrt{1-\rho} W_1^d, \tilde{W}_1^d).$$

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

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

$$F_d := \frac{d^{1/2}}{2\eta} \left(\log \left\| \sqrt{\rho} \tilde{W}_1^d + \sqrt{1-\rho} W_1^d \right\|^2 - \log \|\tilde{W}_1^d\|^2 \right), \tag{4.1}$$

since the law does not change. The other properties of F_d can be studied via 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. *For any absolutely continuous function $f : \mathbb{R} \to \mathbb{R}$,*

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

for some constant $C > 0$.

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

4.2 Some properties of Q_d

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

Lemma 4.3. *For $d \in \mathbb{N}$, let ξ_d follows the chi-squared distribution with d degrees of freedom. For $k \in \mathbb{Z}$ such that for $-2k < d$, there exists $C > 0$ such that*

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

Moreover, for $k > 2$,

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

Proof. The first clam follows from

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

The second part follows from Example A.2. □

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

Proposition 4.4. $\|q_d - q\|_\infty \to 0, \|q_d' - q'\|_\infty \to 0$.

Proof. The probability density q_d is

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

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

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

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

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

and for $x < 2C$,

$$\mathbb{E}\left[\left|q\left(x\frac{d}{\xi_d}\right) - q(x)\right|\right] \le \mathbb{E}\left[\left|q\left(x\frac{d}{\xi_d}\right) - q(x)\right|, \left|\frac{d}{\xi_d} - 1\right| > 1/2\right] + \mathbb{E}\left[\left|q\left(x\frac{d}{\xi_d}\right) - q(x)\right|, \left|\frac{d}{\xi_d} - 1\right| \le 1/2\right]$$
$$\le \|q\|_\infty 4\mathbb{E}\left[\left|\frac{d}{\xi_d} - 1\right|^2\right] + \epsilon \to \epsilon \ (d \to \infty).$$

Thus $\|q_\infty - q\|_\infty \to 0$. 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 3.2

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

$$R^d_m = r_d(X^d_m),\ R^{d*}_m = r_d(X^{d*}_m) \tag{4.2}$$

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

$$Af(y) = d^{-1}(\mathbb{E}[f(R^d_1)|R^d_0 = y] - f(y)).$$

See Section 4.2 of Ethier and Kurtz [1986] for the detail. We will apply Theorem IX.4.21 of Jacod and Shiryaev [2003] to the process $\tilde{Y}^d$. If $\tilde{Y}^d$ converges to a limit, then Y^d converges to the same limit by Lemma B.1.

Proof of Proposition 3.2. For the proof, we need to show some asymptotic properties of conditional distribution of R^d_1 given $R^d_0 = y \in \mathbb{R}$. For notational simplicity, we write y for R^d_0, and write $\mathbb{P}_y$ and $\mathbb{E}_y$ for the conditional probability and expectation given $R^d_0 = y$. By using this notation, we have $R^{d*}_1 = y + \eta d^{-1/2}F_d$. Let

$$\begin{cases} a_d(y) &= d\mathbb{E}_y[R^d_1 - R^d_0], \\ b_d(y) &= d\mathbb{E}_y[(R^d_1 - R^d_0)^2], \\ c_d(y) &= d\mathbb{E}_y[(R^d_1 - R^d_0)^4]. \end{cases} \tag{4.3}$$

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

$$a_d(y) = d\mathbb{E}_y\left[\left(R^{d*}_1 - R^d_0\right)\min\left\{1, \frac{\tilde{q}_d(R^{d*}_1)}{\tilde{q}_d(R^d_0)}\right\}\right] = \eta d^{1/2}\mathbb{E}_y\left[F_d \min\left\{1, \frac{\tilde{q}_d(y + \eta d^{-1/2}F_d)}{\tilde{q}_d(y)}\right\}\right].$$

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

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

by Markov's inequality. Suppose that $\tilde{q}'(y) > 0$. We can find an open bounded neighborhood $\mathcal{O}$ 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 $z \in \mathcal{O}$,

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

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

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

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 $\psi_\epsilon(d^{-1/2}F_d)\to 1$, we have

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

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

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

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

and hence

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

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

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

locally uniformly.

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

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

Then,

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

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

$$c_d(y)=d\mathbb{E}_y\left[\left(R_1^d-R_0^d\right)^4\right]\le\eta^4 d^{-1}\mathbb{E}_y[|F_d|^4]=o(1)\ (d\to\infty).$$

Thus we obtain the convergence of the triplet (4.3). This convergence corresponds to the conditions (i) and (ii) of Theorem IX.4.21 of Jacod and Shiryaev [2003], and the condition (iii) corresponds to $\mathcal{L}(R_0^d)=\tilde{Q}_d\to\tilde{Q}$, 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 Karatzas and Shreve [1991] and hence condition IX.4.3 (i) holds, and measurability follows by Exercise 6.7.4 of Stroock and Varadhan [1979] and hence condition IX.4.3 (ii) holds. Thus the convergence $\tilde{Y}^d \Rightarrow Y$ follows from Theorem IX.4.21 of Jacod and Shiryaev [2003]. 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 Karatzas and Shreve [1991]. Hence the claim follows. □

Proof of Theorem 3.3. 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

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

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,

$$\mathbb{P}\left(|R_{m-1}^d - R_m^d| \ge \epsilon\ \exists m \in \{1,\ldots,M-1\}\right) \le M\mathbb{P}\left(|\eta d^{-1/2}F_d| \ge \epsilon\right) \to 0$$

and

$$\mathbb{P}\left(|\|\tilde{W}_m^d\|^2/d - 1| \ge \epsilon\ \exists m \in \{1,\ldots,M\}\right) \le M\mathbb{P}\left(|\|\tilde{W}_1^d\|^2/d - 1| \ge \epsilon\right) \to 0.$$

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

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

for $m \ge 1$, where $R_0 \sim \tilde{Q}$ and $S_0 \sim N_k(0, \exp(2R_0)I_k)$. By Proposition 3.2, the process $Y^d = \left\{R_{[dt]}^d\right\}_t$ converges to a stationary ergodic process. Hence the claim follows by Lemma B.3. □

4.4 Inconsistency for the pCN algorithm

Lemma 4.5. *Let $\{X_m^d\}_m$ be the Markov chain generated by the pCN algorithm. Then, for any $\epsilon > 0, k \in \mathbb{N}$,*

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

Proof. We have

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

Let

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

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 4.3.

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

$$R_1^{d*} - R_0^d \le \frac{1}{2}\log\left(\rho + 2\sqrt{\rho(1-\rho)}e^{-\epsilon}|d^{-1/2}A_d| + (1-\rho)e^{-2\epsilon}(1 + 2d^{-1/2}B_d)\right) =: \frac{1}{2}\log(1 - \xi + d^{-1/2}C_d)$$

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

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

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

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

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

$$\mathbb{P}(R_0^d > R_1^d > \epsilon) = o(d^{-k}), \mathbb{P}(R_0^d < R_1^d < -\epsilon) = o(d^{-k})$$

and hence

$$\mathbb{P}(|R_0^d| > \epsilon, R_1^d \ne R_0^d) = o(d^{-k}).$$

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

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

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

$$\begin{aligned}\mathbb{P}(|R_0^d| > \epsilon, X_1^d \ne X_0^d) &\le \mathbb{P}(|R_0^d| > \epsilon, X_1^d \ne X_0^d, R_1^d \ne R_0^d) + \mathbb{P}(|R_0^d| > \epsilon, X_1^d \ne X_0^d, R_1^d = R_0^d)\\ &\le \mathbb{P}(|R_0^d| > \epsilon, R_1^d \ne R_0^d) + \mathbb{P}(X_1^d \ne X_0^d, R_1^d = R_0^d) = o(d^{-k}).\end{aligned}$$

□

Lemma 4.6. *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 $\tilde{Q}(\mathcal{O}) \ge \delta$ for some $\delta > 0$. By $\tilde{Q}_d \to \tilde{Q}$ and by Lemma 4.5, for any $p \in \mathbb{N}$,

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

Thus we have the following degenerate property:

$$\liminf_{d\to\infty} \mathbb{P}\left(\frac{1}{d^p}\sum_{m=0}^{d^p-1} f \circ \pi_1(X_m^d) = f \circ \pi_1(X_0^d)\right) \ge \delta$$

for any bounded continuous function $f(x)$ where $\pi_1(x) = x_1$ is the first component of the vector $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

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 $T_d/d^p \to 0$. Then the following should also be satisfied:

$$\frac{1}{d^p}\sum_{m=0}^{d^p-1} f \circ \pi_1(X_m^d) - P_d(f \circ \pi_1) = o_{\mathbb{P}}(1).$$

Note here that since P_d is the scale mixture of the normal distribution, $\mathcal{L}(\pi_1(X_0^d)) = P_1$. Hence we have

$$P_1(\{x; |f(x) - P_1(f)| < \epsilon\}) = \liminf_{d\to\infty} \mathbb{P}(|f \circ \pi_1(X_0^d) - P_d(f \circ \pi_1)| < \epsilon) \geq \delta$$

for any $\epsilon > 0$. 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 $T_d/d^p \to 0$ for any $p > 0$ and hence the pCN algorithm cannot have a polynomial rate of convergence. □

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 two anonymous referees for their extremely careful readings of this manuscript and their many insightful comments which lead to numerous improvements.

A Properties of F_d via Malliavin calculus

A.1 Basic operators in Malliavin calculus

We will study asymptotic properties of F_d defined in (4.1). The basic tool will be Malliavin calculus. The following is a quick review of Malliavin calculus. For the detail, see monographs such as Nualart [2006] and Nourdin and Peccati [2012].

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 $H_n(x) = (-1)^n e^{x^2/2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2/2}/n!$ be the n-th Hermite polynomial:

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

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 $\{H_n(W(h)); h \in \mathfrak{H}, \|h\|_{\mathfrak{H}} = 1\}$. 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 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

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

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

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

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

By the so-called hypercontractivity property of Ornstein-Uhlenbeck operator, we have the following for finite Wiener chaoses. See Corollary 2.8.14 of Nourdin and Peccati [2012] for the proof.

Proposition A.1. *Let $F \in \mathcal{H}_n$. 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* $\|h\|_{\mathfrak{H}} = 1$.

Example A.2. *If ξ_d follows the chi-squared distribution with d degrees of freedom, $\mathbb{E}[(\xi_d/d - 1)^p]^{1/p} \leq (p-1)\sqrt{2/d}$. 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.2 (Theorem 3.3.1 of Nourdin and Peccati [2012])**.**

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

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

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 Malliavin calculus. The connection between Stein's technique and 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 Nourdin and Peccati [2012] for the proof. See also the proof of Theorem 3.1 of Nourdin and Peccati [2009] to replace smoothness of f by the existence of the density of F.

Proposition A.3 (Theorem 2.9.1 of Nourdin and Peccati [2012])**.** *Suppose that $F \in \mathbb{D}^{1,2}$ has a density with respect to the Lebesgue measure. Then for any absolutely continuous function f,*

$$|\mathbb{E}\left[(F - \mathbb{E}\left[F\right]) f(F)\right] - \mathbb{E}\left[f'(F)\right]| \le \|f'\|_\infty \mathbb{E}\left[\left|1 - \langle DF, -DL^{-1}F\rangle\right|\right].$$

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

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

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

Then

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

are in the 2nd Wiener chaos $\mathcal{H}_2$ since

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

We also define

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

which is in $\mathcal{H}_2$ since

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

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

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

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

$$F_d = \frac{d^{1/2}}{2\eta}\left(\log(d + 2A_d) - \log(d + 2B_d)\right).$$

An important remark is that the random variable F_d defined has a density with respect to the Lebesgue measure. In general, a random variable in a finite sum of Wiener chaoses has a density by Theorems 5.1 and

5.2 of Shigekawa [1980]. 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

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

is non-degenerate, $F_d = (d^{1/2}/2\eta)(\log(d+2A_d) - \log(d+2B_d))$ has a density.

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

$$\mathbb{E}\left[\left(\frac{d+2A_d}{d}\right)^k\right] = 1 + O(d^{-1}) \tag{A.4}$$

for any $k \in \mathbb{Z}$ by Lemma 4.3. 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.4.

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

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 $\log 1 + x \le x$,

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

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

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

by Schwartz inequality. The first expectation in the right-hand side is $O(1)$ by Lemma 4.3. For the second expectation, by Proposition A.1 and Minkowski inequality,

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

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

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

Proposition A.5. *Suppose that f is an absolutely continuous function. Then*

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

for some $C > 0$. In particular, $\|\mathcal{L}(F_d) - N(0,1)\|_{\mathrm{TV}} \le 2Cd^{-1/2}$.

Proof. We will apply Proposition A.3 by the estimate of $1 - \langle DF_d, -DL^{-1}F_d\rangle_{\mathfrak{H}}$. Since F_d is not in a finite Wiener chaoses, we consider an approximation

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

where A_d and B_d are defined in (A.3). We show that this approximation is valid in the sense that

$$d^{1/2}\mathbb{E}\left[\|DF_d - DF_{d,0}\|_{\mathfrak{H}}^2\right]^{1/2} = O(1). \tag{A.5}$$

Observe that

$$DF_{d,0} = \frac{d^{-1/2}}{2\eta}(DA_d - DB_d)$$

and

$$DF_d = \frac{d^{1/2}}{2\eta}\left(\frac{DA_d}{d+2A_d} - \frac{DB_d}{d+2B_d}\right) = DF_{d,0} + \frac{d^{-1/2}}{2\eta}DA_d\left(\frac{d}{d+2A_d} - 1\right) - \frac{d^{-1/2}}{2\eta}DB_d\left(\frac{d}{d+2B_d} - 1\right).$$

Therefore, since (A_d, DA_d) and (B_d, DB_d) have the same law, we have

$$\begin{aligned} d^{1/2}\mathbb{E}\left[\|DF_d - DF_{d,0}\|_{\mathfrak{H}}^2\right]^{1/2} =& \frac{1}{2\eta}\mathbb{E}\left[\left\|DA_d\left(\frac{d}{d+2A_d} - 1\right) - DB_d\left(\frac{d}{d+2B_d} - 1\right)\right\|_{\mathfrak{H}}^2\right]^{1/2} \\ \leq& \frac{1}{\eta}\mathbb{E}\left[\left\|DA_d\left(\frac{d}{d+2A_d} - 1\right)\right\|_{\mathfrak{H}}^2\right]^{1/2}. \end{aligned}$$

Since $\|DA_d\|_{\mathfrak{H}}^2 = d + 2A_d$, the integrand in the right-hand side is

$$(d+2A_d)\left(\frac{d}{d+2A_d} - 1\right)^2 = d\left\{\left(\frac{d}{d+2A_d} - 1\right) + \left(\frac{d+2A_d}{d} - 1\right)\right\}$$

and hence the expectation is $O(1)$ by (A.4). Thus, approximation $F_{d,0}$ is valid in the sense of (A.5).

The approximated variable $F_{d,0}$ has much simpler structure than that of F_d. We have

$$\|DF_{d,0}\|_{\mathfrak{H}}^2 = \frac{d^{-1}}{4\eta^2}\left(\|DA_d\|_{\mathfrak{H}}^2 - 2\langle DA_d, DB_d\rangle_{\mathfrak{H}} + \|DB_d\|_{\mathfrak{H}}^2\right) = \frac{d^{-1}}{4\eta^2}((d+2A_d) - 2(\sqrt{\rho}C_d + d\rho) + (d+2B_d)),$$

and hence

$$\mathbb{E}\left[\|DF_{d,0}\|_{\mathfrak{H}}^2\right] = \frac{d^{-1}}{4\eta^2}(2d - 2d\rho) = O(1).$$

By this estimate together with Lemma A.4, we have $F_d \in \mathbb{D}^{1,2}$. Since $F_{d,0} \in \mathcal{H}_2$, we have $-L^{-1}F_{d,0} = F_{d,0}/2$, and hence $-DL^{-1}F_{d,0} = DF_{d,0}/2$. Therefore,

$$\left\langle DF_{d,0}, -DL^{-1}F_{d,0}\right\rangle_{\mathfrak{H}} = \frac{1}{2}\|DF_{d,0}\|_{\mathfrak{H}}^2$$

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

$$\begin{aligned} \mathbb{E}\left[\left|1 - \langle DF_{d,0}, -DL^{-1}F_{d,0}\rangle_{\mathfrak{H}}\right|\right] &= \mathbb{E}\left[\left|1 - \frac{d^{-1}}{8\eta^2}\left(2d - 2d\rho + 2A_d - 2\sqrt{\rho}C_d + 2B_d\right)\right|\right] \\ &= \frac{d^{-1}}{8\eta^2}\mathbb{E}\left[\left|2A_d - 2\sqrt{\rho}C_d + 2B_d\right|\right] = O(d^{-1/2}). \end{aligned}$$

The last step of this proof is to show

$$\mathbb{E}\left[\left|1 - \langle DF_d, -DL^{-1}F_d\rangle_{\mathfrak{H}}\right|\right] = O(d^{-1/2}). \tag{A.6}$$

If the above holds, we have the claim by Proposition A.3. First, observe that by triangular inequality,

$$\begin{aligned}\mathbb{E}\left[\left|1-\langle DF_d,-DL^{-1}F_d\rangle_{\mathfrak{H}}\right|\right] \leq & \mathbb{E}\left[\left|1-\langle DF_{d,0},-DL^{-1}F_{d,0}\rangle_{\mathfrak{H}}\right|\right] \\ & + \mathbb{E}\left[\left|\langle DF_d,-DL^{-1}F_d\rangle_{\mathfrak{H}}-\langle DF_{d,0},-DL^{-1}F_{d,0}\rangle_{\mathfrak{H}}\right|\right]\end{aligned}$$

Thus the proof will be finished when the second term is $O(d^{-1/2})$. Again by the triangular inequality and Hölder's inequality, the second term is bounded above by

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

where in the second inequality, we used $\mathbb{E}[\|DL^{-1}F\|_{\mathfrak{H}}^2]\leq\mathbb{E}[\|DF\|_{\mathfrak{H}}^2]$. Hence we have (A.6) which is sufficient for our claim. □

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

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

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

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

Then $N^d(\tau^d(t))=[dt]$ and hence $Y_t^d=\tilde{Y}_{\tau^d}^d$. Therefore, by Proposition VI.6.37 of Jacod and Shiryaev [2003], it is sufficient to show

$$\lim_{d\to\infty}\mathbb{P}\left(\sup_{s\leq S}|\tau_s^d-s|>\epsilon\right)=0$$

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

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

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

$$N^d(\max\{s-\epsilon,0\})\geq N^d(\tau_s^d)=[ds]\geq d\max\{s-\epsilon,0\}.$$

Thus, for $S'=S+\epsilon$, the above probability is bounded by

$$\mathbb{P}\left(\sup_{s\leq S'}\left|\frac{N_s^d}{d}-s\right|\geq\epsilon\right)\leq\epsilon^{-2}\mathbb{E}\left[\sup_{s\leq S'}\left|\frac{N_s^d}{d}-s\right|^2\right]\leq 4\epsilon^{-2}\mathbb{E}\left[\left|\frac{N_{S'}^d-S'}{d}\right|^2\right]=4\epsilon^{-2}\frac{dS'}{d^2}=o(1),$$

by 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 Kamatani [2014a]). *Let $X^d = \{X^d_m\}_m$ 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*

$$\frac{1}{M}\sum_{m=0}^{M-1} f(X^d_m) - \mathbb{E}[f(X^d_0)] = o_{\mathbb{P}}(1)\ (M, d \to \infty)$$

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

Proof. Since $\mathcal{L}(X^d_0) \to \mathcal{L}(X_0)$, we can substitute $\mathbb{E}[f(X^d_0)]$ by $\mathbb{E}[f(X_0)]$ in the above equation, and hence it is sufficient to show

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

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

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

Then, by stationarity,

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

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^d_m has two parts, $X^d_m = (X^{d,1}_m, X^{d,2}_m)$ where $X^{d,i}_m$ 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

$$P_d(\mathrm{d}x_1\mathrm{d}x_2) = P^1_d(\mathrm{d}x_1)P^2_d(\mathrm{d}x_2|x_1).$$

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

Assumption 2.

1. *For $Y^d_t = X^{d,1}_{[T_dt]}$, $Y^d \Rightarrow Y$ (in Skorohod's sense) where Y is stationary and ergodic continuous process with the invariant probability measure P^1.*
2. *Random variables $X^d = \{X^d_m\}_m$ converges to $X = \{X_m\}_m = \{(\xi, X^2_m)\}_m$ where $\xi \sim P^1$ and conditioned on ξ, the process $X^2 = \{X^2_m\}_m$ is stationary and ergodic with the invariant probability measure $P^{2|1}(\cdot|\xi)$.*
3. *For any bounded continuous function f, $P^{2|1}f(x_1) = \int f(x_1, x_2)P^{2|1}(\mathrm{d}x_2|x_1)$ is continuous in x_1.*

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

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

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

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

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

for f such that $\int f(x)P^{2|1}(\mathrm{d}x_2|x_1)P(\mathrm{d}x_1) = 0$. For such f and for $\epsilon > 0$, choose M_0 so that

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

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

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{1}{T_d M}\sum_{m=0}^{T_d M-1} g(X^{d,1}_m)\right|\right] &= \mathbb{E}\left[\left|\frac{1}{M}\int_0^{M} g(Y^d_t)\mathrm{d}t\right|\right]\\
&\le \mathbb{E}\left[\left|\frac{1}{M}\sum_{k=0}^{[M/M_0]-1}\int_{kM_0}^{(k+1)M_0} g(Y^d_t)\mathrm{d}t\right|\right] + \mathbb{E}\left[\left|\frac{1}{M}\sum_{k=0}^{[M/M_0]-1}\int_{[M/M_0]M_0}^{M} g(Y^d_t)\mathrm{d}t\right|\right]\\
&\le \frac{M_0}{M}\left[\frac{M}{M_0}\right]\mathbb{E}\left[\left|\frac{1}{M_0}\int_0^{M_0} g(Y^d_t)\mathrm{d}t\right|\right] + \mathbb{E}\left[\left|\frac{1}{M}\sum_{k=0}^{[M/M_0]-1}\int_{[M/M_0]M_0}^{M} g(Y^d_t)\mathrm{d}t\right|\right]\\
&\to \mathbb{E}\left[\left|\frac{1}{M_0}\int_0^{M_0} g(Y_t)\mathrm{d}t\right|\right] \le \epsilon/2.
\end{aligned}$$

Set $S_d = T_d M$. We still need to show

$$\limsup_{d,M\to\infty}\mathbb{E}\left[\left|\frac{1}{S_d}\sum_{m=0}^{S_d-1} f(X^d_m) - \frac{1}{S_d}\sum_{m=0}^{S_d-1} g(X^{d,1}_m)\right|\right] \le \epsilon/2.$$

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

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

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

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{1}{S_d}\sum_{m=0}^{S_d-1} f(X^d_m)\right|\right] &\le \frac{S_0}{S_d}\left[\frac{S_d}{S_0}\right]\mathbb{E}\left[\left|\frac{1}{S_0}\sum_{m=0}^{S_0-1} f(X^d_m)\right|\right] + \|f\|_\infty\frac{S_d - [S_d/S_0]S_0}{S_d}\\
&\to \mathbb{E}\left[\left|\frac{1}{S_0}\sum_{m=0}^{S_0-1} f(X_m)\right|\right] \le \epsilon/2.
\end{aligned}$$

Thus, we can conclude that

$$\limsup_{d,M\to\infty} \mathbb{E}\left[\left|\frac{1}{T_dM}\sum_{m=0}^{T_dM-1} f(X_m^d)\right|\right] \le \limsup_{d,M\to\infty}\left\{\mathbb{E}\left[\left|\frac{1}{T_dM}\sum_{m=0}^{T_dM-1} g(X_m^{d,1})\right|\right] + \mathbb{E}\left[\left|\frac{1}{T_dM}\sum_{m=0}^{T_dM-1}(f(X_m^d)-g(X_m^{d,1}))\right|\right]\right\}$$
$$\le \epsilon.$$

Hence the proof is completed. □

References

D. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. *J. Roy. Statist. Soc. Ser. B*, 36:99–102, 1974. ISSN 0035-9246.

Alexandros Beskos, Gareth Roberts, and Andrew Stuart. Optimal scalings for local metropolis-hastings chains on nonproduct targets in high dimensions. *Ann. Appl. Probab.*, 19(3):863–898, 06 2009. doi: 10.1214/08-AAP563.

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

S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, 28(3):424–446, 2013. ISSN 0883-4237. doi: 10.1214/13-STS421.

Andreas Eberle. Error bounds for Metropolis-Hastings algorithms applied to perturbations of Gaussian measures in high dimensions. *Ann. Appl. Probab.*, 24(1):337–377, 2014. ISSN 1050-5164. doi: 10.1214/13-AAP926.

Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. ISBN 0-471-08186-8. doi: 10.1002/9780470316658. Characterization and convergence.

Charles J. Geyer. Practical markov chain monte carlo. *Statist. Sci.*, 7(4):473–483, 11 1992. doi: 10.1214/ss/1177011137.

Fumitaka Goto. An extension and practical performances of mpcn algorithm. Master's thesis, Graduate School of Engineering Science, Osaka University, 2017.

Martin Hairer, Andrew M. Stuart, and Sebastian J. Vollmer. Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6):2455–2490, 2014. ISSN 1050-5164. doi: 10.1214/13-AAP982.

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

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

Kengo Kamatani. Rate optimality of Random walk Metropolis algorithm in high-dimension with heavy-tailed target distribution. *Arxiv*, 2014b.

Kengo Kamatani. Ergodicity of Markov chain Monte Carlo with reversible proposal. *Journal of Applied Probability*, 54, 2017.

Kengo Kamatani and Masayuki Uchida. Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Statistical Inference for Stochastic Processes*, 18(2):177–204, 2015. ISSN 1387-0874. doi: 10.1007/s11203-014-9107-4.

Kengo Kamatani, Akihiro Nogita, and Masayuki Uchida. Hybrid multi-step estimation of the volatility for stochastic regression models. *Bulletin of Informatics and Cybernetics*, March 2016.

Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus.* Number 113 in Graduate texts in mathematics. Springer-Verlag, 2nd ed edition, 1991.

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

Radford M. Neal. Regression and classification using Gaussian process priors. In *Bayesian statistics, 6 (Alcoceber, 1998)*, pages 475–501. Oxford Univ. Press, New York, 1999.

Ivan Nourdin and Giovanni Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 145 (1-2):75–118, 2009. ISSN 0178-8051. doi: 10.1007/s00440-008-0162-x.

Ivan Nourdin and Giovanni Peccati. *Normal approximations with Malliavin calculus. From Stein's method to universality*, volume 192 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 2012. ISBN 978-1-107-01777-1. doi: 10.1017/CBO9781139084659.

Ivan Nourdin and Guillaume Poly. Convergence in total variation on Wiener chaos. *Stochastic Process. Appl.*, 123(2):651–674, 2013. ISSN 0304-4149. doi: 10.1016/j.spa.2012.10.004.

David Nualart. *The Malliavin calculus and related topics.* Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006. ISBN 978-3-540-28328-7; 3-540-28328-5.

Natesh S. Pillai, Andrew M. Stuart, and Alexandre H. Thiery. Optimal Proposal Design for Random Walk Type Metropolis Algorithms with Gaussian Random Field Priors. *Arxiv*, 2014.

Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11, 2006.

Christian P. Robert and George Casella. *Monte Carlo statistical methods.* Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004. ISBN 0-387-21239-6. doi: 10.1007/978-1-4757-4145-2.

Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling of discrete approximations to langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(1):255–268, 1998. ISSN 1467-9868. doi: 10.1111/1467-9868.00123.

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

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

Ichiro Shigekawa. Derivatives of Wiener functionals and absolute continuity of induced measures. *J. Math. Kyoto Univ.*, 20(2):263–289, 1980. ISSN 0023-608X.

D.W. Stroock and S.R.S. Varadhan. *Multidimensional Diffussion Processes.* Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete. Springer Verlag, 1979. ISBN 9783540903536.

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