On the analysis of some recursive equations in Probability

Arunangshu Biswas

Indian Statistical Institute, Kolkata

September, 2014

On the analysis of some recursive equations in Probability

Arunangshu Biswas

Thesis submitted to the Indian Statistical Institute

in partial fulfilment of the requirements

for the award of the degree of

Doctor of Philosophy

September, 2014

Indian Statistical Institute

203, B.T. Road, Kolkata, India.

"*The greatest and most important problems of life are all in a certain sense insoluble. They can never be solved, but only outgrown* "

- Carl Jung, 1875 - 1961, psychiatrist, psychoanalyst.

Acknowledgements

First and foremost I want to thank my supervisor Prof. G. K. Basak, Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, whose knowledge and support has constantly guided me in the course towards the completion of this work.

I am thankful to my family who have always encouraged me in this endeavour.

Special thanks is also due to all the faculty in the Stat. Math. Unit (SMU) and Applied Statistics Unit (ASU) from whom I have taken many interesting courses in Probability and Statistics. I am grateful to all of them.

I am indebted to my colleagues and staff members of the Department of Statistics, Presidency University (formerly College) where I have been associated for the last few years. I would like to take this opportunity to thank all of them.

Contents

Dedicated to my mother...

Chapter 1

Introduction

1.1 Prelude

This thesis deals with recursive systems used in theoretical and applied probability. Recursive systems are stochastic processes $\{X_n\}_{n\geq 1}$ where the X_n depends on the earlier X_{n-1} and also on some increment process which is uncorrelated with the process X_n. The simplest example of a recursive system is the Random Walk, whose properties have been extensively studied. Mathematically a recursive system takes the form

$$X_n = f(X_{n-1}, \epsilon_n),$$

ϵ is the increment/ innovation procedure and $f(\cdot,\cdot)$ is a function on the product space of x_n and ϵ_n.

We first consider a recursive system called Self-Normalized sums (SNS) corresponding to a sequence of random variables $\{X_n\}$ (which is assumed to be symmetric about zero). Here the sum of X_i is normalized by an estimate of the p^{th} absolute moment constructed from the X_i's. The SNS are the most conservative among all normalized sums in the sense that all the moments of the SNS exist even if X_i do not possess any finite moments. We look at the functional version of the SNS called the Self-Normalized Process (SNP) where the X_i's come from a very general family called the domain of attraction of the stable distribution with stability index α denoted by $DA(\alpha)$, for $\alpha \in (0,2]$ (for definition see Section 2.2). We show that for any choice of α and p other than 2 the limiting distributions of the SNP are either trivial or do not exist.

We consider another recursive system called the Adaptive Markov Chain Monte Carlo (AMCMC) which is used extensively in statistical simulation. The motivation behind this method is to get hold of a Markov Chain (MC) whose stationary distribution (if it exists) is the distribution of interest, also called the target distribution, henceforth denoted as $\psi(\cdot)$. One chooses a proposal distribution which is a conditional probability distribution, say $p(\cdot|x)$ and then given a present value of the chain at x_n generates a new value $y \sim p(\cdot|x_n)$. The new value y is accepted with a certain probability, called acceptance probability, which depends on the target distribution. It can be verified that the MC constructed in this way has $\psi(\cdot)$ as the stationary distribution.

The usual choice of the proposal given the value of x_n is a distribution which is symmetric about the mean x_n, say for example, Normal with mean 0 and variance σ^2. Therefore one has :

$$X_{n+1} = X_n + \epsilon I(U < \alpha_n), \text{ where } \epsilon \sim N(0, \sigma^2),$$

and U is an Uniform $(0, 1)$ random variable and $\alpha_n = \min\{1, \frac{\psi(X_n+\epsilon)}{\psi(X_n)}\}$.

The problem with this choice is that even though in the long run this process X_n may converge to $\psi(\cdot)$ the convergence may be show for bad choices of σ^2. In practice the choice of the unknown parameters that determine the speed of convergence are made to depend on the present and/or past values of the chain in addition to some additional quantities. This is called Adaptive MCMC (AMCMC) in statistical literature. We deal with such an MC where the parameters depend on the present and/or past values of the chain and on an indicator variable which takes the value one if the last generated sample was accepted. It is not certain a priori that such an MC will also have $\psi(\cdot)$ as its invariant distribution. One aspect of this thesis is to explore the convergence criteria of such adaptive chains along with the their rate of convergence. We apply the diffusion approximation procedure, which is basically scaling down the discrete process to a continuous time diffusion process governed by a Stochastic Differential Equation (SDE). The gain is that not only can we invoke standard convergence results for diffusion process, but we can also apply the same diffusion approximation to other suitably defined processes and then apply various techniques to compare their relative efficiency. This is possible since there exists many discretization schemes for diffusion processes. See Kloeden and Platen [39] for example.

Although the standard Normal distribution is the standard choice of the proposal distribution we investigate whether other choices of the proposal and the various target distributions also yield similar diffusion approximation results. We classify target distributions according to three classes, each corresponding to some condition that ensures the existence or non existence of the m.g.f of the target density $\psi(\cdot)$. We also show that this condition is not necessary by explicitly considering the standard Cauchy as the target density. We further prove a Theorem that rules out the heavy tailed distribution, in particular the standard Cauchy distribution as a candidate for the proposal distribution. This type of diffusion approximation is the cumulative addition of increments from the proposal distribution normalized by a quantity θ which also changes with each iteration. It is therefore possible to look upon it as a version of normalized sums of $X_i's$ where the normalization is by θ and the $X_i's$ comes from the proposal distribution. In this context we connect some of the results of the SNP to the diffusion approximation of AMCMC.

1.2 Self-Normalized Processes

The first example of recursive equation is what are popularly called the Self-Normalized Process (SNP). This topic is dealt in Chapter 2. To understand the SNP we first define the Self-Normalized Sums (SNS):

$$Y_{n,2} = S_n/V_{n,2}; \quad \text{where} \quad S_n = \sum_{i=1}^{n} X_i; \quad V_{n,2} = (\sum_{i=1}^{n} X_i^2)^{\frac{1}{2}},$$

under the assumptions that the denominator is never zero almost surely. This can be written as a recursive process in $Y_{n,2}$:

$$Y_{n+1,2} = S_{n+1}/V_{n+1,2} = \Big(\frac{V_{n,2}}{V_{n+1,2}} Y_{n,2} + \frac{X_{n+1}}{V_{n+1,2}}\Big).$$

The origin of the study of the SNS is the Students t statistic which dates back to 1908 when William Gosset ("Student") [28] considered the problem of statistical inference on the mean μ when the standard deviation σ is unknown. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from a distribution $F(\cdot)$ and let $\overline{X}_n = n^{-1} \sum\limits_{i=1}^{n} X_i$ be the sample mean and

$s_n^2 = (n-1)^{-1}\sum_{i=1}^{n}\Big(X_i - \overline{X}_n\Big)^2$ be the sample variance. The t- statistics is then defined as

$$T_n = \sqrt{n}\Big(\frac{\overline{X}_n - \mu}{s_n}\Big).$$

If $F(\cdot)$ is the $N(\mu, \sigma^2)$ distribution then above statistics follows the t distribution with $n-1$ degrees of freedom, which tends to the standard Normal distribution as $n \to \infty$. It has been shown that the limiting distribution of T_n is Normal distribution even if $X_i's$ do not follow Normal distribution, see [40]. When non-parametric tests were subsequently introduced the t statistics was compared to non-parametric tests (sign test, rank test, etc.) where $F(\cdot)$ typically had 'fat' tails with infinite second moment or even first absolute moments.

Observe that when $\mu = 0$ the t- statistics and the SNS are related by:

$$T_n = Y_{n,2}\Big\{\frac{n-1}{n - Y_{n,2}^2}\Big\}^{1/2}.$$

From the above relation, if T_n or $Y_{n,2}$ has a limiting distribution then so does the other and they coincide, see Proposition (1) of Griffin [31]. Efron [20] and Logan B.F., Mallows C. L., Rice S. O. and Shepp, L. A.,[40] derived the asymptotic distribution of SNS. Active development in the Self-Normalized sums began in the late 1990's with the seminal works of Griffin and Kuelbs [29, 30] on laws of iterated logarithms for SNS of i.i.d. variables belonging to the domain of attraction of a Normal or a stable law. Subsequently Bentkus and Götze [10] derived a Berry Essen bound for the t statistics. The interest in the asymptotics of SNS was renewed in the seminal work of Giné E., Götze F. and Mason M. S. [27] who characterized the convergence of the SNS as:

$$\frac{S_n}{V_{n,2}} \xrightarrow{\mathcal{L}} N(0,1)$$

iff

$$E(X_i) = 0 \;\text{ and } X_i \text{ lies in the domain of attraction of a Normal distribution.}$$

Later in Csörgő, Syszkowicz and Wang [17], a related result was proved for the Self-Normalized partial sums processes, $S_{[nt]}/V_{n,2},\ 0 \leq t \leq 1$, namely

$$\frac{S_{[nt]}}{V_{n,2}} \xrightarrow{\mathcal{L}} W(t) \text{ in } D[0,1],$$

iff $X_i's$ follow the same conditions as in Giné *et al.* [27] (viz., $E(X_i) = 0$ and X_i lies in the domain of attraction of the Normal distribution), where $W(\cdot)$ is the Brownian motion in $[0,1]$ and $D[0,1]$ is the space of all cádlág functions in $[0,1]$. Since weak convergence in [0,1] implies convergence of the finite dimensional distributions, the necessary condition of this result comes from the results of Giné *et al* [27]. An important contribution of the paper by Csörgő *et al* [17] is the fact that the Self-Normalized version of the Donsker's invariance principle also holds in the domain of attraction of the Normal law, even without assuming the finite variance. This was in spirit with some properties of the SNS where many standard results like the LIL and moderate deviations hold with much less assumptions on the finiteness of the moments (see Griffin and Kuelbs [29] and Shao [61]). The natural choice of the normalizing variable in the denominator of the SNS is the L_2 normalization given by $(\sum_{i=1}^{n} X_i^2)^{1/2}$. We investigate whether it is possible to find other modes of normalization. Again it is clear that the normalization has something to do with the index of stability parameter $\alpha \in (0,2]$ of the ingredient random variables X_i. So basically we have an infinite number of combinations of the parameter α and normalization parameter $p > 0$. It is then certainly meaningful to ask for what choices of (α, p) do we get a non-trivial limit. The same questions can be asked for the process version of the SNS, called the Self-Normalized Process (SNP) defined later.

In a related (unpublished) work Basak and Dasgupta [4] showed the convergence of a suitably scaled SNP to an Ornstein-Uhlenbeck (OU) process. There again, the ingredient random variables come from the domain of attraction of a Normal distribution (stable distribution with $\alpha = 2$) and the normalization is L_2 normalization. That work motivated us to ask whether similar results can be guaranteed for other random variables and other choices of normalization.

In this thesis we are concerned with the process version of the Self-Normalized sums given by :

$$Y_{n,p}(t) = \frac{S_{[nt]}}{V_{n,p}} + (nt - [nt])\frac{X_{[nt]+1}}{V_{n,p}} \qquad 0 < t < 1 \quad p > 0, \tag{1.2.1}$$

where

$$V_{n,p}^p = \sum_{i=1}^{n} |X_i|^p.$$

This process $Y_{n,p}(\cdot)$ is quite the same as that of Csörgő *et al.* [17] except that we make it continuous by interpolating it between each sub intervals. In Chapter 2 we will look into all possible combinations of the pair (p, α) and by the elimination process find the pairs that give a nontrivial (i.e. non-degenerate) limiting distribution for the process.

1.3 MCMC and the Metropolis Hastings (MH) Algorithm

The second example of recursive scheme that we consider is the Adaptive MCMC. But before that we define what are commonly called the Markov Chain Monte Carlo (MCMC) procedures. MCMC techniques have gained huge recognition over the last two decades. It has been increasingly applied to diverse fields such as computer sciences, finance, meteorology, statistical genetics and many others. It is used very much by Bayesians, to simulate from the posterior distribution for general choice of prior distributions when the normalizing constant is unknown. Its applicability can be gauged from the fact that only twenty years since inception, a whole handbook of around six hundred pages has been dedicated to MCMC theory and methods (see the book edited by S. Brooks, A. Gelman, G. Jones, X. L. Meng [15]).

1.3.1 MCMC

Quite often one is required to find the integral of a complicated function (possibly multidimensional), say $\int f(\mathbf{x})d\mathbf{x}$. Assume that the standard techniques of numerical integration, e.g. Gauss quadrature, are not easily usable in this case. Also assume that the integral can be written in an equivalent way as:

$$\int f(\mathbf{x})d\mathbf{x} = \int g(\mathbf{x})\psi(\mathbf{x})d\mathbf{x} := E_{\psi}[(g(\mathbf{X})],$$

where $\psi(\cdot)$ is a density function , i.e., $\psi(\cdot) \geq 0$, $\int \psi(\mathbf{x})d\mathbf{x} = 1$. The Monte Carlo solution to this problem is to generate a sample of size n, say $X_1, X_2, \ldots, X_n$ from the distribution whose density function is $\psi(\cdot)$ and then approximate

$$\int f(\mathbf{x})d\mathbf{x} \approx \frac{1}{n}\sum_{i=1}^{n} g(X_i).$$

The above approximation is valid since by the Strong Law of large Numbers (SLLN) we have that if $Var_{\psi}(g(X)) < \infty$ then

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \stackrel{a.s}{\rightarrow} E_{\psi}(X),$$

which is the required integral. Also from the Central Limit Theorem, one can approximate the error by

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} g(X_i) - E_{\psi}(g(X))\Big) \stackrel{d}{\rightarrow} N(0, V_{\psi}(g(X))).$$

Therefore the computation of the above integral boils down to the generation of a sample from a distribution $\psi(\cdot)$ henceforth called the *target* density. The standard techniques of simulation is the Inverse Transform method, which requires the existence of the inverse of the distribution function in a closed form, the Accept- Reject algorithm, which requires that the target distribution be dominated by a density from which sample generation is easy, or any other transformation method. Consequently, the applicability of such methods are slightly limited.

MCMC methods tackle the problem of simulation from a different perspective. Suppose one agrees to sacrifice precision for a solution then he is willing to go for a sample that is not exactly generated from $\psi(\cdot)$ but approximately from $\psi(\cdot)$. The MCMC methods entails the existence of an aperiodic, irreducible Markov chain $\{X_n\}$ on $\mathcal{X}$, which is the support of the target distribution and whose invariant distribution is $\psi(\cdot)$. Under these conditions it is guaranteed that the distribution of X_n will tend to $\psi(\cdot)$ as $n \to \infty$. In fact a stronger statement is true

$$||P^n(x,\cdot) - \psi(\cdot)||_{TV} \to \infty$$

where the total variation (TV) of a measure $\nu(\cdot)$ on $(\Omega, \mathcal{F})$ is defined as $||\nu(\cdot)||_{TV} = \sup_{A\in\mathcal{F}} \nu(A)$ and $P^n(x,\cdot) = P(X_n \in \cdot | X_0 = x)$, see [55]. TV convergence implies distributional convergence. Therefore in practice one would choose a very large n_0 called *burn in* and look at the sample $\{X_{n_0}, X_{n_0+1}, X_{n_0+2}, \ldots X_{n_0+m}\}$ of size m. This sample although not identical or independent has distribution very similar to $\psi(\cdot)$. Therefore the problem boils down to : Given a target distribution $\psi(\cdot)$ how to obtain an aperiodic, irreducible Markov Chain whose invariant distribution is $\psi(\cdot)$. This proverbial needle in a haystack problem has a very simple solution given by the Metropolis- Hastings (MH) algorithm and the Gibbs sampler.

1.3.2 MH algorithm

The MH algorithm, originally proposed by Metropolis *et al.* [44] and introduced in statistical contexts by Hastings [32], constructs such a Markov chain in a surprisingly simple way. Let $\psi(\cdot)$ have a (possibly un-normalized) density, say ψ_u. Let $P(x,\cdot)$ be any other Markov chain whose transitions also have a (possibly un-normalized) density, i.e., $Q(x,dy) \propto q(x,y)dy$. Then this method proceeds as follows. First choose some X_0. Then, given X_n, generate a *proposal* Y_{n+1} from $Q(X_n,\cdot)$. Also flip an independent coin, whose probability of heads equals $\alpha(X_n, Y_{n+1})$, where

$$\alpha(x,y) \quad = \quad \min\Big[1, \frac{\psi_u(y)q(y,x)}{\psi_u(x)q(x,y)}\Big],$$

where $q(x,y)$ is the density of $Q(x,\cdot)$ (assuming it exists). This choice for the acceptance probability is due to the following definition and proposition.

Definition: A measure $\psi(\cdot)$ satisfies the detailed balance (reversibility) condition if

$$\int_E \psi(dx)P(x,F) = \int_F \psi(dy)P(y,E),$$

for some sets E and F belonging to the Borel σ-algebra defined on the state space of the Markov chain.
A consequence of reversibility is :

PROPOSITION 1. If a Markov chain is reversible with respect to $\psi(\cdot)$, then $\psi(\cdot)$ is the stationary distribution of the chain.

Proof: We compute that for some set $A \in \sigma(\mathcal{X})$

$$\int_{\mathcal{X}} \psi(dx)P(x,A) = \int_A \psi(dy)P(y,\mathcal{X}) = \int_A \psi(dy) = \psi(A),$$

since $P(y,\mathcal{X}) = 1, \forall y \in \mathcal{X}$. This proves stationary. ■

PROPOSITION 2. The MH algorithm described above satisfies the detailed balance condition with respect to $\psi(\cdot)$.

Proof: We need to show

$$\psi(dx)P(x,dy) = \psi(dy)P(y,dx),$$

where $P(x,dy) = q(x,y)\alpha(x,y)dy$. It suffices to assume that $x \neq y$ (since if $x = y$ then the equations are trivial). But for $x \neq y$, setting $c = \int_{\mathcal{X}} \psi_u(x)dx$,

$$\begin{aligned}
\psi(dx)P(x,dy) &= [c^{-1}\psi_u(x)dx][q(x,y)\alpha(x,y)dy] \\
&= c^{-1}\psi_u(x)q(x,y)\min\{1, \frac{\psi_u(y)q(y,x)}{\psi_u(x)q(x,y)}\}dxdy \\
&= c^{-1}\min[\psi_u(x)q(x,y), \psi_u(y)q(y,x)]dxdy
\end{aligned}$$

which is symmetric in x and y. ■

Therefore to run the Metropolis Hastings algorithm on a computer, we just need to be able to run the proposal chain $Q(x,\cdot)$ (which is easy for some suitable choices, say $Normal(0,\sigma_x^2)$) and to compute the acceptance probabilities and then do the accept/

reject steps. Furthermore we only need to compute the ratios of densities, so we don't require the Normalizing constant c.
This method is very liberal on the choice for $Q(x, \cdot)$. In fact depending on the different forms for $Q(x, \cdot)$ we have different versions of the MH algorithm, see, for example [55], such as:

- **Symmetric Metropolis Algorithm.** Here $q(x, y) = q(y, x)$ and hence $\alpha(x, y) = \min\{1, \frac{\psi_u(y)}{\psi_u(x)}\}$.
- **Random Walk MH**. Here $q(x, y) = q(y - x)$. For example, perhaps $Q(x, \cdot) = N(x, \sigma^2)$, or $Q(x, \cdot) = Uniform(x - 1, x + 1)$.
- **Independence Sampler**. Here $q(x, y) = q(y)$
- **Langevin algorithm**: Here the proposal is generated by

$$Y_{n+1} \sim N(X_n + \delta/2 \bigtriangledown \log(\psi(X_n)), \delta) \quad \delta > 0.$$

The last form is motivated by the discrete approximation to Langevin diffusion.

In this thesis we will be concerned with the Symmetric Random Walk MH (SRW MH) algorithm (defined in the next subsection).

1.3.3 Drawbacks of the MH algorithm and the AMCMC procedure

1.3.3.1 The optimal scaling problem in the RW MH algorithm

Let ψ_u be the un-normalized target distribution. Consider running an MH algorithm for ψ_u. The optimal scaling problem concerns the question of how should we choose the proposal density for this algorithm. For example consider the RW MH algorithm with the proposal distribution given by $Q(x, \cdot) = N(x, \sigma^2)$. In this case $\alpha = \min\{1, \frac{\psi(y)}{\psi(x)}\}$. This is also referred to as the Normal Symmetric RW MH (N-SRW MH) algorithm. Then the problem becomes how should one choose σ.

If σ^2 is chosen to be very small then the proposed move y will be very near to the present value x of the chain. Since α is the ratio of the densities at x and y, it will be very close to 1. Consequently, the new point will have a high probability of acceptance. Now, since the new value will be close to the chain's previous value the chain will move extremely slowly, leading to a very high acceptance rate even if the current point is in the valley of $\psi(\cdot)$ distribution, thus yielding a very poor performance. On the other hand, if σ^2 is chosen to be too large, then the proposed value will usually be very far from the current state. Consequently, the acceptance probability $\alpha(x, y)$ is likely to be very close to zero if the current state is near a peak of the $\psi(\cdot)$ distribution. Unless the chain gets very lucky the proposed value will almost never be accepted and the chain will get 'stuck' at the same state for a long time with poor acceptance rates. Thus, the choice of the proposal scaling σ^2 should therefore be 'just right' (also called the Goldilocks principle after J. Rosenthal, see [56]).

In a paper by Gelman A., Roberts G. and Gelman W. [25], the authors provided a partial answer to this problem. Their method is outlined as follows:

To start let us assume that the un-normalized target distribution for the multivariate $\mathbf{X} = (X^{(1)}, X^{(2)}, \ldots, X^{(d)}) \in \mathbb{R}^d$, for some $d \in \mathbb{N}$, is :

$$\psi_u(\mathbf{x}) = \prod_{i=1}^{d} f(x_i), \tag{1.3.2}$$

for $\mathbf{x} \in \mathbb{R}^d$ and $x_i \in \mathbb{R}$, i.e., the target density is the product of the marginals. Although this case is simple it provides useful insights which can be approximated in other cases as well. For the RW-MH with multivariate Normal proposals set the proposal variance to be $\sigma_d^2 \mathbf{I}_d$ where $\sigma_d^2 = \frac{l^2}{d}$ where l is a constant to be chosen later. Let $\{\mathbf{X}_n\} = \{(X_n^{(1)}, X_n^{(2)}, \ldots, X_n^{(d)})\}$ be the MC on $\mathbb{R}^d$. Also let $\{N(t)\}_{t\geq 0}$ be a Poisson process with rate d which is independent of $\{X_n\}$. Finally let

$$Z_t^d = X_{N(t)}^{(1)}, \ t \geq 0.$$

Using results from Ethier and Kurtz [22] it has been shown in Gelman *et al.* [25] that as $d \to \infty$ the process $\{Z_t^d\}_{t\geq 0}$ converges weakly to the diffusion process $\{Z_t\}_{t\geq 0}$ which satisfies the SDE

$$dZ_t = h(l)^{1/2} dB_t + \frac{1}{2} h(l) \bigtriangledown \log \psi_u(Z_t) dt.$$

Here

$$h(l) = 2l^2\Phi(-\frac{\sqrt{I}l}{2})$$

where $I = E((\log f)'(Z))^2$, Z having density $f(\cdot)$. Here $h(\cdot)$ corresponds to the speed of the limiting diffusion. Numerically it turns out that the speed measure is maximized if $l = \hat{l} = 2.38/\sqrt{I}$. If $f(\cdot)$ is the density of the $N(0,1)$ distribution, then I is 1 which implies that the optimal value of l is 2.38. Also it is proved that the optimal asymptotic acceptance rate of the algorithm is 0.234. So this method prescribes an optimal value of the acceptance ratio albeit under a restrictive scenario. The conditions on the form of the target distribution (1.3.2) was later extended by Roberts and Rosenthal [54] who extended the above results for non-homogeneous target densities of the form

$$\psi(\mathbf{x}) = \prod_{i=1}^{d} C_i f(C_i x_i),$$

where C_i are real numbers, and later by Bedard and Rosenthal [8] who considered the case of target distributions $\psi(\cdot)$ which had independent coordinates with vastly different scaling.

1.3.3.2 Adaptive MCMC

The method in the previous sections only outlines what can be the possible acceptance rates for an MH algorithm, but does not suggest any method by which that optimality can be reached. For example, if the optimal acceptance rate for the independent MH (i.e., the MH whose target is of the form (1.3.2)) is 0.234, the question is how one can find the appropriate proposal scaling whose optimal acceptance is as above.

One naive solution to this problem is to hand tune the algorithm to reach the optimal acceptance rates. So for example, if the empirical acceptance rate exceeds the optimal level one can decrease the proposal scaling by a quantity δ_1; if the empirical acceptance rate is below the optimal acceptance rate, one can similarly increase the proposal scaling by a quantity, say δ_2. However this method has a drawback in the sense that it requires a manual supervisor who would monitor the output of the chain. Also the choices of δ would then be subjective and the outputs cannot be obtained on a real time basis.

An alternative to this technique is to construct an algorithm that tunes the proposal 'on the fly', i.e., on a real time basis. Mathematically, let $\{P_\gamma\}_{\gamma \in \mathcal{Y}}$ be a set of proposal kernels indexed by the scale γ, also called the *adaptation* parameter, in an adaptation set $\mathcal{Y}$, each of which has $\psi(\cdot)$ as its stationary distribution, i.e., $(\psi P_\gamma)(\cdot) = \psi(\cdot)$. Assuming that P_γ is ϕ- irreducible and aperiodic (which it usually will be), this implies (see, for exmple, Meyn and Tweedie [45]) that P_γ will be *ergodic* for $\psi(\cdot)$, that is

$$||P_\gamma(x, \cdot) - \psi(\cdot)||_{TV} \to 0, \tag{1.3.3}$$

for all $x \in \mathcal{X}$, see [55] for the definition. Here $||\mu(\cdot)||_{TV}$ is the total variation norm of a measure defined by $||\mu(\cdot)||_{TV} = \sup_{A \in \mathcal{F}} |\mu(A)|$. Let Γ_n be the adaptation parameter at the n^{th} iteration of the algorithm. Therefore the proposal kernel will be given by $P_{\Gamma_n}(\cdot, \cdot)$. The updation at this iteration given the value of $X_n = x$ and $\Gamma_n = \gamma$ will be governed by the probability

$$P(X_{n+1} \in A | X_n = x, \Gamma_n = \gamma, X_{n-1}, \ldots, X_0, \Gamma_{n-1}, \ldots, \Gamma_0) = P_\gamma(x, A), \tag{1.3.4}$$

for $n = 0, 1, 2, \ldots$. Similarly Γ_n are updated according to some updating algorithm. In principle the choice of Γ_n can be made to depend on the infinite past, though in practice it is often the case that the pair $\{X_n, \Gamma_n\}$ is a Markov chain. However since the aim of the algorithms is to generate a sample from $\psi(\cdot)$ it is not straightforward that the chain will preserve ergodicity. Ergodicity, for the above process (1.3.4), is defined in Roberts and Rosenthal [57]) as,

$$T(\mathbf{x}, \gamma, n) := ||A^{(n)}((\mathbf{x}, \gamma), \cdot) - \psi(\cdot)||_{TV} \to 0, \text{ as } n \to \infty, \tag{1.3.5}$$

where

$$A^{(n)}((\mathbf{x}, \gamma), B) = P(X_n \in B | X_0 = \mathbf{x}, \Gamma_0 = \gamma).$$

This means that whatever be the starting point (x, γ), the chain $\{X_n\}$ always converges in the TV norm to $\psi(\cdot)$. In general, any process $\{X_n\}$ will not necessarily be ergodic even if for every fixed $\gamma \in \mathcal{Y}$ the kernel P_γ is stationary. This is shown by a counter example in Chapter 4 of [15] which we reproduce below.

Example 1: Let $\mathcal{Y} = \{1, 2\}$, let $\mathcal{X} = \{1, 2, 3, 4\}$, let $\psi(1) = \psi(3) = \psi(4) = 0.333$ and $\psi(2) = 0.001$. Let each P_γ be an RW MH algorithm, with proposal $Y_{n+1} \sim \mathcal{U}\{X_n - 1, X_n + 1\}$ for P_1, or $Y_{n+1} \sim \mathcal{U}\{X_n - 2, X_n - 1, X_n + 1, X_n + 2\}$ for P_2. Define the adaptation by letting $\Gamma_{n+1} = 2$ if the n^{th} proposal was accepted, otherwise $\Gamma_{n+1} = 1$. Then each P_γ is reversible with respect to ψ, since, for example,

$$\psi(1)P_1(1,2) = \psi(1)\frac{1}{2}\min\{1, \frac{\psi(2)}{\psi(1)}\} = \frac{1}{2}\psi(2) = \psi(2)\frac{1}{2}\min\{1, \frac{\psi(1)}{\psi(2)}\} = \psi(2)P_1(2,1).$$

Similarly it can be shown for other values of (x, y, γ) that

$$\psi(x)P_\gamma(x,y) = \psi(y)P_\gamma(y,x).$$

However since $\frac{\psi(2)}{\psi(1)} = \frac{1}{333}$ the chain can get stuck at $X_n = \Gamma_n = 1$ for a long period of time. Hence the limiting distributions will be skewed heavily toward 1 and less towards 3 and 4.

This example clearly shows that naive algorithms are not necessarily ergodic. ■

Roberts and Rosenthal [57] gave a set of sufficient conditions for which an adaptive algorithm will be ergodic. Their conditions are:

PROPOSITION 3. 1. **Simultaneous Uniform Ergodicity condition**: For every $\epsilon > 0$ there exists $N_0 = N_0(\epsilon) \in \mathbb{N}$ such that $\forall N \geq N_0$,

$$||P_\gamma^N(\mathbf{x}, \cdot) - \psi(\cdot)||_{TV} \leq \epsilon,$$

for every fixed $\mathbf{x} \in \mathcal{X}$ and $\gamma \in \mathcal{Y}$, where $P_\gamma^N(\mathbf{x}, \cdot) = P_\gamma(X_N \in \cdot | X_0 = \mathbf{x})$ and,

2. **Diminishing Adaptation condition**: $\lim_{n\to\infty} D_n = 0$ in probability, where

$$D_n = \sup_{x\in\mathcal{X}} ||P_{\Gamma_{n+1}}(\mathbf{x}, \cdot) - P_{\Gamma_n}(\mathbf{x}, \cdot)||_{TV},$$

is a $\mathcal{G}_{n+1}$ random variable where

$$\mathcal{G}_n = \sigma\{X_0, X_1, \ldots, X_n, \Gamma_0, \Gamma_1, \ldots, \Gamma_n\}.$$

The first condition says that the time to convergence to ergodicity should be uniformly bounded over all the adaptation parameters $y \in \mathcal{Y}$ and starting points $x \in \mathcal{X}$; the second condition says that the change in the transition kernel over each iteration (measured in the sense of total variation norm) should tend to zero as n tends to infinity. It has also been pointed out in the same paper that it is not required that the sum of the adaptation be finite, in other words one can possibly have $\sum_{n=1}^{\infty} D_n = \infty$ with probability one.

It is easy to construct chains that satisfy condition (2). For example, one way to incorporate this into an algorithm is to change the transition kernel at the $(n+1)^{st}$ iteration with probability $p(n)$, such that $p(n) \to 0$ as $n \to \infty$.

Using the results of Gelman *et al.* [25], Haario, Saksman and Tamminen [33] were the first to suggest the following Adaptive MH scheme. Their algorithm ran as :

Algorithm 1

1. Start with an initial $\mathbf{X}_0 \in \mathbb{R}^d$.
2. At the time $n-1$ one has sampled $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{n-1}$. Choose a candidate point $\mathbf{Y} \in \mathbb{R}^d$ from the proposal distribution $q_n(\cdot|\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{n-1}) \sim N(\mathbf{X}_{n-1}, C_n)$ where

$$C_n = \begin{cases} C_0 & n \leq n_0 \\ s_d Cov(\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{n-1}) + s_d \epsilon \mathbf{I}_d & n > n_0 \end{cases}$$

   where $n_0, s_d, \epsilon > 0$ are suitably chosen constants and C_0 is a suitably chosen dispersion matrix.

From the results of Gelman *et al.* [25] one choice of the parameter s_d is $2.38^2/d$, see Haario *et al.* [33] for more details. Harrio *et al.* [33] also came up with an ergodic result for the adaptive chain described here(Algorithm 1):

Proposition 4. (Theorem (1) of Haario *et al.* [33]): Let $\psi(\cdot)$ be the density of a target distribution supported on a bounded measurable subset $\mathcal{S} \subset \mathbb{R}^d$, and assume that $\psi(\cdot)$ is bounded from above. Let $\epsilon > 0$ and μ_0 be any initial distribution on S. Then the above chain simulates properly the target distribution $\psi(\cdot)$: For any bounded and measurable function $f : S \to \mathbb{R}$ the equality

$$\lim_{n\to\infty} \frac{1}{n+1}(f(X_0) + f(X_1) + \cdots + f(X_n)) = \int_{\mathcal{S}} f(x)\psi(dx),$$

holds almost surely.

1.3.3.3 An adaptive MCMC based on the empirical acceptance rates

Based on the discussions above we suggest an adaptive MCMC (described here for the univariate case only) that is based on the empirical acceptance rate. A slight variation of the present algorithm was suggested by Prof Peter Green via personal communication :
Algorithm 2

1. Start with an initial $(\mathbf{X}_0, \theta_0, \xi_0) \in (\mathbf{X} \times (0, \infty) \times \{0, 1\})$, where $\mathbf{X}$ is the state space.

2. At time $n-1$ one has sampled (X_i, θ_i, ξ_i) for $i = 1(1)n-1$. Propose a new point from a Normal distribution, i.e., $Y \sim N(X_{n-1}, \theta_{n-1})$.

3. Accept the new point with the MH acceptance probability $\alpha(X_{n-1}, Y) = \min\{1, \frac{\psi(Y)}{\psi(X_{n-1})}\}$. If the point is accepted, set $X_n = Y$, otherwise $X_n = X_{n-1}$.

4. Update
$$\theta_n = \theta_{n-1} e^{\frac{1}{\sqrt{n}}(\xi_n - q)}, \text{ where } q > 0 \Leftrightarrow \log(\theta_n) = \log(\theta_{n-1}) + \frac{1}{\sqrt{n}}(\xi_n - q)$$

5. Increase n by one unit and repeat the above from step 2.

Let us describe the algorithm. θ_n is the proposal scaling (tuning) parameter which is adaptively tuned depending on whether the previous sample was accepted. If the generated sample Y is accepted then $\theta_n > \theta_{n-1}$, (i.e., $\xi_n = 1$) thus increasing the proposal variance at the next step, allowing the chain to explore more regions in the state space. If the sample was rejected (i.e., $\xi_n = 0$) then $\theta_n < \theta_{n-1}$, thus making the next proposal move slightly conservative. Here q is a benchmark value. From the discussions in Section 1.3.3.1, a value for q for Normal target with independent components was suggested by Gelman *et al.* [25] to be 0.234. This algorithm, in principle is similar to the Stochastic Approximation procedure. See Andrieu and Moulines [1] for the connection between the adaptive MCMC and the stochastic approximation procedure.
Chapters 3 and 4 will be devoted towards proving the asymptotic results about this chain using the diffusion approximation procedure. In Chapter 5 we relax the assumption on the target and proposal distribution used in the previous two chapters and also give a

brief description of the diffusion approximation applied to multivariate AMCMC and its limiting distribution.

Chapter 2

Self-Normalized Processes

2.1 Self-Normalized sums as recursive equations

The first example of recursive equations that we investigate is what is popularly called Self-Normalized sums (SNS). Let $\{X_i\}$ be a sequence of i.i.d random variables from the distribution $F(\cdot)$. Then the SNS corresponding to $\{X_n\}$ is defined as

$$Y_{n,p} = S_n/V_{n,p}$$

where

$$S_n = \sum_{i=1}^{n} X_i; \quad V_{n,p}^p = \sum_{i=1}^{n} |X_i|^p.$$

This can be looked upon as a recursive system by re-writing $Y_{n,p}$ as

$$Y_{n+1,p} = \frac{V_{n,p}}{V_{n+1,p}} Y_{n,p} + \frac{X_{n+1}}{V_{n+1,p}}.$$

In this chapter we will investigate the functional version of the SNS defined as:

$$Y_{n,p}(t) := \frac{S_{[nt]}}{V_{n,p}} + (nt - [nt]) \frac{X_{[nt]+1}}{V_{n,p}},$$

where $[x]$ is the greatest integer less than or equal to x, for any $x \in \mathbb{R}$.

2.2 Basic facts about Stable distributions

Stable distributions are a class of distributions that share some common property. The need for stable distribution arose from the fact that for distributions with no finite variance (for example Cauchy) the CLT does not hold with $\sqrt{n}$ normalization. However properly normalized (which is n for Cauchy) the sequence of partial sums converge to a distribution (in the above case to a Cauchy distribution). The family of limiting distribution comprises the class of stable distributions. A formal definition is given below. In this section we follow the convention due to Samorodnitsky and Taqqu [59].

Definition 1: Stable definition

A random variable X is said to have a stable distribution if it has a *domain of attraction*, i.e., if there exists a sequence of i.i.d random variables $\{Y_n\}$ and sequences of reals $\{a_n\}$ and positive reals $\{d_n\}$ such that :

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{d_n} + a_n \Rightarrow X.$$

Some other equivalent definitions are given as following:

Definition 2: A random variable is said to have a stable distribution if for any positive numbers A and B, there is a positive number C and a real number D such that

$$AX_1 + BX_2 \stackrel{d}{=} CX + D,$$

where X_1, X_2 are independent copies of X. The following proposition can be proved, see Feller [23], Section VI.1 for a proof.

PROPOSITION 5. For any stable distribution X, there is a number $\alpha \in (0, 2]$ such that the constant C in the above definition satisfies

$$C^{\alpha} = A^{\alpha} + B^{\alpha}.$$

The number α is called the *index of stability*, *characteristic exponent* or *stability index*. A stable random variable X with index α is called α- stable. For example, if X follows $N(\mu, \sigma^2)$ then the following is true:

$$AX_1 + BX_2 \sim N((A + B)\mu, (A^2 + B^2)\sigma^2)$$

this implies

$$C^2 = A^2 + B^2, \text{ and } D = (A + B - C)\mu.$$

Therefore X has a stable distribution with $\alpha = 2$.
From the classical CLT the Normal distribution is a stable distribution with stability index $\alpha = 2$. The totality of the domain of attraction of Stable distribution with index $(\alpha)(:= S(\alpha))$ is denoted by $DA(\alpha)$. Also, the totality of all distributions belonging to the domain of attraction of a Normal distribution is denoted by DAN.

Another equivalent definition is

Definition 3: If $X_1, X_2, \ldots, X_n$ are i.i.d. copies of X then

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} C_n X + D_n.$$

Remark 1. It turns out (see Feller [23], Theorem VI.1.1) that necessarily $C_n = n^{1/\alpha}$ where α is defined in the previous definition. In Definition (1) if X is the Normal distribution then all distributions having finite variance belong to the domain of attraction of the Normal law by the statement of the ordinary central limit Theorem. Y_i's are said to belong to the domain of attraction of $S(\alpha)$ if $d_n = n^{1/\alpha} h(n)$ where $h(x),\ x > 0$ is a slowly varying function (at infinity), i.e., $\lim\limits_{x\to\infty} h(ux)/h(x) = 1,$ for all $u > 0$.

We state a property of slowly varying function. For a proof see, for example Senata [60] Theorem 1.1.

Proposition 6. **Uniform Convergence Theorem:** If $L(\cdot)$ is a slowly varying function, then for every fixed $[a, b], 0 < a < b < \infty$ the relationship

$$\lim_{x\to\infty} \frac{L(\lambda x)}{L(x)} = 1$$

holds uniformly with respect to $\lambda \in [a, b]$.

Definition 4: Another definition of stable laws is by characterization through characteristic functions. In fact, it is the most useful definition since none of the stable distributions except the Levy $(\alpha = 0.5)$, Cauchy$(\alpha = 1)$ and Normal $(\alpha = 2)$ admits a closed form of the density. However in this thesis we are not going to use that characterization.

We list the following properties of the Stable distributions and for distributions belonging to the domain of attraction of $S(\alpha)$. For a proof see Feller [23];

1. For $X \in DA(\alpha)$, $\alpha \in (0,2)$, $E|X|^p < \infty$, $\forall\ 0 < p < \alpha$. If $X \in DA(2)$ then the second moment may also be finite. If $X \in S(2)$ all the moments are finite.

2. If $X \in DA(\alpha)$, $P(|X| > x) = x^{-\alpha}h(x)$, $\forall x > 0$, where $h(\cdot)$ is a slowly varying function on $[0,\infty)$.

We state the Karamata's Theorem for slowly varying functions that will be required later. For a proof see, for example, [21], Theorem A3.6.

PROPOSITION 7. **Karamata's Theorem**: If $h(\cdot)$ is a slowly varying function then

$$\frac{1}{h(x)}\int_{x_0}^{x}\frac{h(t)}{t}dt \to \infty, \text{ as } x \to \infty, \text{ for some } x_0 > 0.$$

We also state a result due to Lemma 3.2 of Giné *et al* [27]:

PROPOSITION 8. If $X_i's$ are i.i.d and belong to $DA(2)$, $E(X_i) = 0$ and S_n and V_n are defined as earlier then

$$\begin{aligned}\frac{S_n}{\sqrt{nl(n)}} &\to N(0,1) \text{ in distribution}\\ \frac{V_{n,2}^2}{nl(n)} &\to 1 \text{ in probability.}\end{aligned}$$

for some function $l(n)$ which is slowly varying at ∞. In case of finite variance $l(n) = E(X_i^2) < \infty, \forall n$.

The following is a characterization of DAN:

LEMMA 1. If X_i are i.i.d and and X_1 are symmetric about zero then

$$E\Big(\frac{X_1^4}{(\sum_{i=1}^{n} X_i^2)^2}\Big) = o(\frac{1}{n}) \Leftrightarrow X_i \in DAN.$$

Proof: The only if part is proved in Theorem 3.3 (see Equation 3.7) of Giné *et al.* [27]. For the if part, from part (a) of the Theorem

$$F \in DAN,\ E(X_1) = 0 \Rightarrow Y_n := \frac{\sum_{i=1}^{n} X_i}{(\sum_{i=1}^{n} X_i^2)^2} \xrightarrow{d} N(0,1).$$

Since Y_n converges in distribution it is stochastically bounded. Hence by Corollary 2.6 (or Remark 2.7) of Gine *et al.* [27] there is convergence of moments to the moments of the limiting distribution. Consequently $\lim_{n\to\infty} E(\frac{S_n}{V_{n,2}})^4 = 3$. Moreover from Equation (3.8) of [27]

$$\begin{aligned} E(\frac{S_n}{V_{n,2}})^4 &= 3 - 2nE(\frac{X_1}{V_{n,2}})^4 + 8\binom{n}{2}E(\frac{X_1X_2^3}{V_{n,2}^4}) + 36\binom{n}{3}E(\frac{X_1X_2X_3^2}{V_{n,2}^4}) \\ &+ 24\binom{n}{4}E(\frac{X_1\ldots X_4}{V_{n,2}^4}). \end{aligned} \tag{2.2.1}$$

Following arguments similar Equation (2.4.5) and (2.4.6) we have that $E(\frac{X_1}{V_{n,2}^4}|X_2,\ldots,X_n)$ is zero. Hence the third, fourth and fifth expectation in the RHS of Equation (2.2.1) are zero. Taking limits as $n \to \infty$ on both sides of Equation (2.2.1) we have that $nE(\frac{X_1}{V_{n,2}})^4 \to 0$ which proves that $E(\frac{X_1}{V_{n,2}})^4 = o(\frac{1}{n})$. ■.

We state a result, due to Darling [19], Theorem 5.1, on the limiting distribution of $\frac{\sum_{i=1}^{n} X_i}{X_n^*}$ where $X_n^* = \max\{X_1, X_2, \ldots, X_n\}$ and $X_i \in DA(\alpha)$ with $0 < \alpha < 1$.

Proposition 9. Let $X_i \geq 0$ and $X_i \in DA(\alpha)$ with $0 < \alpha < 1$, then

$$\lim_{n\to\infty} P\Big(S_n < yX_n^*\Big) = G(y),$$

where $G(y)$ has the characteristic function

$$\int e^{ity} dG(y) = \frac{e^{it}}{1 - \alpha\int_0^1 (e^{itx} - 1)\frac{dx}{x^{\alpha+1}}}.$$

2.3 Self-Normalized sums and processes and the main Theorem

A Self-Normalized sums (SNS) with i.i.d. components is defined as

$$Y_{n,2} = \frac{S_n}{V_{n,2}}, \quad \text{where } S_n = \sum_{j=1}^{n} X_j, \quad V_{n,2}^2 = \sum_{j=1}^{n} X_i^2,$$

where X_i are i.i.d. copies of a random variable X. Recursively,

$$\begin{aligned} Y_{n+1,2} &= \frac{S_n}{V_{n+1,2}} + \frac{X_{n+1}}{V_{n+1,2}} \\ &= \Big(\frac{V_{n,2}}{V_{n+1,2}}\Big) Y_{n,2} + \frac{X_{n+1}}{V_{n+1,2}}. \end{aligned}$$

In fact, it is related to the Students t distribution, and it has been shown in Griffin [31] that the latter has the same asymptotic distribution as the former. The first results in SNS were proved by Efron [20], Logan *et al.* [40] where the latter showed that the asymptotic distribution of the SNS was Normal if X belongs to the domain of attraction of a Normal distribution, i.e., $X \in DAN$. Logan *et al.*. [40] also conjectured the 'only if' part that was proved by Giné *et al.* in 1997 [27] thus renewing an interest in this topic which was followed by works of many authors, see [17, 53, 61].

Extending the works of Giné *et al.* [27], Csörgő *et al.* [17] (also Račkauskas and Sequet [53]) asked whether an invariance formula in the lines of classical FCLT can also be asked for Self-Normalized processes. The answer was in the affirmative which was proved by both of the authors. In Cšorgő *et al.* [17] a functional convergence form of the above result was proved :

PROPOSITION 10. As $n \to \infty$, the following statements are equivalent:

1. $E(X) = 0$ and X is in the domain of attraction of a Normal law.

2. $S_{[nt_0]}/V_{n,2} \xrightarrow{\mathcal{L}} N(0, t_0)$ for $t_0 \in (0, 1]$.

3. $S_{[nt]}/V_{n,2} \xrightarrow{\mathcal{L}} W(t)$ on $(D[0, 1], \rho)$ where ρ is the sup-norm metric.

4. On an appropriate probability space for $X, X_1, X_2, \ldots$ we can construct a standard Wiener Process $\{W(t), 0 \le t < \infty\}$ such that

$$\sup_{0 \le t \le 1} |S_{[nt]}/V_{n,2} - W(nt)/\sqrt{n}| = o_p(1).$$

However in all the works cited above the normalization of the SNS (SNP) is with index 2. A pertinent question is whether we can expect to get similar results when different normalization are taken. It is also clear that in that case we should also vary the choice of the index parameter α of the Stable distribution, so in the general case $\alpha \in (0, 2]$. Contrary to the discontinuous process in Csörgő *et al.* [17] we consider a continuous process $Y_{n,p}(\cdot)$ defined as :

$$Y_{n,p}(t) = \frac{S_{[nt]}}{V_{n,p}} + (nt - [nt])\frac{X_{[nt]+1}}{V_{n,p}}\ ,\ 0 < t < 1,\ p > 0, \tag{2.3.2}$$

where $S_n = \sum_{i=1}^{n} X_i,\ V_{n,p} = \left(\sum_{i=1}^{n} X_i^p\right)^{1/p}$ and the $X_i \in DA(\alpha),\ 0 < \alpha \le 2$. Here is the main Theorem of this chapter:

THEOREM 1. Let X_i be i.i.d copies of a random variable X which is symmetric about zero and $X \in DA(\alpha)$. Let $Y_{n,p}(t),\ 0 < t < 1$, be defined as in (2.3.2). Then $Y_{n,p}(\cdot)$ converges weakly to a Brownian motion if and only if $p = \alpha = 2$.

Proof: The proof is done by the method of elimination. We apply the Prohorov's Theorem (see, for example, Billingsley, [14]): A sequence of probability measure $\{P_n\}$ in $C[0, 1]$ converges if it is tight and the finite dimensional distributions (f.d.d) converge. In Lemmas 5 - 7 we show that the finite dimensional distribution of $Y_{n,p}(t)$ converges to the zero vector if $0 < p < \alpha < 2$ and $0 < p = \alpha < 2$ for any $n \ge 1$. For $p > \alpha$ we obtain in Lemma 8 the limiting form of the characteristic function of the finite dimensional vector which turns out to be non-degenerate. In Lemma 2 we show that the process is tight if and only if $0 < p \le \alpha \le 2$. The only case where we both have finite dimensional convergence and tightness is when $p = \alpha = 2$. The limiting distribution for the SNS in this case was identified by Giné *et al.* as normal. The convergence in sup norm metric for this choice of p and α follows directly from Proposition 10. ■

2.4 Proof of Theorem 1

2.4.1 A preliminary lemma

We first state a characterization result of the $DA(2)$ distributions, see Feller [23] for a proof :

PROPOSITION 11. Let X_i be i.i.d. random variables with distribution function F. In order that there exist constant $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$ such that

$$\frac{\sum_{i=1}^{n} Y_i - b_n}{a_n} \xrightarrow{\mathcal{L}} N(0,1),$$

or, in other words, for $X \in DA(2)$, a necessary and sufficient condition is :

$$\lim_{x\to\infty} \frac{x^2 P(|X_1| > x)}{E(X_1^2 I(|X_i| \leq x))} = 0.$$

Using the above proposition we prove the following lemma :

LEMMA 2. If $X \in DA(\alpha)$ then $Y := sgn(X)|X|^{\alpha/2} \in DAN$, where $sgn(x)$ is the sign function defined by:

$$sgn(x) = \begin{cases} -1, & \text{if } x < 0 \\ 0, & \text{if } x = 0 \\ 1, & \text{if } x > 0. \end{cases}$$

Proof: From Proposition 11 it is necessary and sufficient to prove that:

$$\lim_{y\to\infty} \frac{y^2 P(|Y| > y)}{E(Y^2 I(|Y| < y))} = 0.,$$

Now,

$$\begin{aligned}
y^2P(|Y|>y) &= y^2P(|X|^{\frac{\alpha}{2}}>y)\\
&= y^2P(|X|>y^{\frac{2}{\alpha}})\\
&= y^2h(y^{\frac{2}{\alpha}})(y^{\frac{2}{\alpha}})^{-\alpha}=h(y^{\frac{2}{\alpha}})\\
&\quad \text{by Property (2) of Stable distributions.}
\end{aligned}$$

And,

$$\begin{aligned}
E(Y^2I(|Y|<y)) &= E(|X|^{\alpha}I(|X|^{\frac{\alpha}{2}}\le y))\\
&= E(|X|^{\alpha}I(|X|\le y^{\frac{2}{\alpha}}))\\
&= \int_0^{y^{\frac{2}{\alpha}}} z^{\alpha}dF_{|X|}(z)\\
&= \int_0^{y^{\frac{2}{\alpha}}} (\int_0^{z}\alpha t^{\alpha-1}dt)dF_{|X|}(z).
\end{aligned}$$

Applying Fubini's Theorem and interchanging the order of integration we get

$$\begin{aligned}
E(Y^2I(|Y|<y)) &= \int_0^{y^{\frac{2}{\alpha}}}\alpha\int_t^{y^{\frac{2}{\alpha}}} dF_{|X|}(z)t^{\alpha-1}dt\\
&= \alpha\int_0^{y^{\frac{2}{\alpha}}} P(t<|X|\le y^{\frac{2}{\alpha}})t^{\alpha-1}dt\\
&= \alpha\int_0^{y^{\frac{2}{\alpha}}} P(|X|>t)t^{\alpha-1}dt-\alpha\int_0^{y^{\frac{2}{\alpha}}} P(|X|>y^{\frac{2}{\alpha}})t^{\alpha-1}dt\\
&= \alpha\int_0^{y^{\frac{2}{\alpha}}}\frac{h(t)}{t}dt-h(y^{\frac{2}{\alpha}}).
\end{aligned}$$

Hence,

$$\lim_{y\to\infty}\frac{y^2P(|Y|>y)}{E(Y^2I(|Y|<y))} = 1/\Big(\alpha\lim_{y\to\infty}\frac{1}{h(y^{\frac{2}{\alpha}})}\int_0^{y^{\frac{2}{\alpha}}}\frac{h(t)}{t}dt-1\Big)=0.$$

by Karamata's Theorem, see Proposition 7. ■

2.4.2 Finite Dimensional Convergence

Fix $k \geq 1$. Select $0 < t_1 < t_2 < \ldots < t_k \leq 1$. We look at the finite dimensional distribution of the random vector $\mathbf{Y_{n,k}} = \Big(Y_{n,p}(t_1), Y_{n,p}(t_2), \ldots, Y_{n,p}(t_k)\big)$ as $n \to \infty$. We do this for $p < \alpha,\ p = \alpha$ and $p > \alpha$.

2.4.2.1 Case 1: $p < \alpha$

LEMMA 3. If $p < \alpha,\ \ \frac{S_n}{V_{n,p}} \xrightarrow{P} 0$

Proof: Since $X_i \in DA(\alpha),\ E|X_i|^p < \infty,\ \forall\ 0 < p < \alpha$. Now, $\frac{V_{n,p}}{n^{\frac{1}{p}}} = \Big(\frac{\sum_{i=1}^{n}|X_i|^p}{n}\Big)^{\frac{1}{p}} \overset{SLLN}{\to} \Big(E(|X|^p)\Big)^{\frac{1}{p}} = C < \infty$. Again since $X_i \in DA(\alpha),\ S_n/(n^{1/\alpha}h(n))$ converges in distribution to an $S(\alpha)$ distribution, where $h(\cdot)$ is a slowly varying function at ∞. Therefore

$$\frac{S_n}{V_{n,p}} \;=\; \frac{S_n/n^{\frac{1}{p}}}{V_n/n^{\frac{1}{p}}} = n^{\frac{1}{\alpha}-\frac{1}{p}}h(n)\frac{S_n/(n^{\frac{1}{\alpha}}h(n))}{V_{n,p}/n^{\frac{1}{p}}}$$

Now since $p < \alpha, \frac{1}{\alpha} - \frac{1}{p} < 0$. $h(\cdot)$ is a slowly varying function (whose growth rate is less than polynomial). Therefore $\frac{h(n)}{n^{\frac{1}{p}-\frac{1}{\alpha}}} \to 0$. The ratio $\frac{S_n/(n^{\frac{1}{\alpha}}h(n))}{V_{n,p}/n^{\frac{1}{p}}}$ converges to $S(\alpha)$ in distribution by the Slutsky's Theorem and therefore $S_n/V_{n,p}$ converges to 0 in probability.
■

2.4.2.2 Case 2: $p = \alpha$.

We have the following inequality involving $V_{n,\alpha}$:

LEMMA 4.

$$V_{n,\alpha} \geq V_{n,1} \geq V_{n,\beta} \geq V_{n,2},\ \text{if}\ \ \alpha \leq 1 \leq \beta \leq 2.$$

Proof. For $\alpha \leq 1$, we have

$$\begin{aligned}
\Big(\frac{|x_i|}{\sum |x_i|}\Big)^{\alpha} &\geq \frac{|x_i|}{\sum\limits_{i=1}^{n} |x_i|} \ \forall\ i = 1, 2, \ldots, n, \\
\Rightarrow |x_i|^{\alpha} &\geq \frac{|x_i|}{\sum\limits_{i=1}^{n} |x_i|} (\sum_{i=1}^{n} |x_i|)^{\alpha} \\
\Rightarrow \sum_{i=1}^{n} |x_i|^{\alpha} &\geq (\sum_{i=1}^{n} |x_i|)^{\alpha}.
\end{aligned}$$

The reverse is true for $2 \geq \beta \geq 1$, i.e.,

$$(\sum_{i=1}^{n} |x_i|)^{\beta} \geq \sum_{i=1}^{n} |x_i|^{\beta}.$$

Combining the two we have

$$\begin{aligned}
(\sum_{i=1}^{n} |x_i|^{\alpha})^{\frac{\beta}{\alpha}} &\geq (\sum_{i=1}^{n} |x_i|)^{\beta} \geq \sum_{i=1}^{n} |x_i|^{\beta} \\
\Rightarrow (\sum_{i=1}^{n} |x_i|^{\alpha})^{\frac{1}{\alpha}} &\geq (\sum |x_i|^{\beta})^{\frac{1}{\beta}}.
\end{aligned}$$

First take $\alpha \leq 1$ and $\beta = 1$ and then $\alpha = 1$ and $\beta \geq 1$ to get,

$$\begin{aligned}
(\sum_{i=1}^{n} |x_i|^{\alpha})^{\frac{1}{\alpha}} \geq \sum |x_i| \geq (\sum_{i=1}^{n} |x_i|^{\beta})^{\frac{1}{\beta}} \\
\Rightarrow V_{n,\alpha} \geq V_{n,1} \geq V_{n,\beta} \qquad (2.4.3)
\end{aligned}$$

Again consider the inequality

$$(\sum_{i=1}^{n} |y_i|)^{p} \geq \sum_{i=1}^{n} |y_i|^{p} \text{ for } p > 1.$$

Applying the above with $y_i = |x_i|^\beta$ and $p = 2/\beta \in [1,2]$ we get:

$$
\begin{aligned}
(\sum_{i=1}^n |x_i|^\beta)^{\frac{2}{\beta}} \geq \sum_{i=1}^n |x_i|^{\beta(2/\beta)} &= \sum |x_i|^2 \Rightarrow (\sum_{i=1}^n |x_i|^\beta)^{\frac{1}{\beta}} \geq (\sum_{i=1}^n |x_i|^2)^{\frac{1}{2}} \\
&\Rightarrow V_{n,\beta} \geq V_{n,2}. \qquad (2.4.4)
\end{aligned}
$$

Combining (2.4.3) and (2.4.4) we have the lemma. ■

LEMMA 5. If $0 < p = \alpha \leq 1$ and X_i are symmetric about 0, then $\lim_{n\to\infty} E(\frac{S_n}{V_{n,p}})^2 = 0$.

Proof Note that,

$$
\begin{aligned}
E\Big(\frac{\sum_{i=1}^n X_i}{V_{n,\alpha}}\Big)^2 &= \sum_{i=1}^n E\Big(\frac{X_i^2}{V_{n,\alpha}^2}\Big) + \sum_{(i,j):i\neq j} E\Big(\frac{X_i X_j}{V_{n,\alpha}^2}\Big) \\
&= \sum_{i=1}^n E\Big(\frac{X_i^2}{V_{n,\alpha}^2}\Big) + \sum_{i=1}^n E\Big(\sum_{j\neq i} \frac{X_i}{V_{n,\alpha}} E\Big(\frac{X_j}{V_{n,\alpha}} \mid X_i, i \neq j\Big)\Big)
\end{aligned}
\qquad (2.4.5)
$$

For the second term note that since X_i are symmetric about 0, which implies that :

$$
\frac{X_j}{V_{n,\alpha}} \stackrel{d}{=} -\frac{X_j}{V_{n,\alpha}}. \qquad (2.4.6)
$$

Also from Lemma 4 we have

$$
\begin{aligned}
\frac{V_{n,1}}{V_{n,\alpha}} \leq 1 \quad &\Rightarrow \quad \sum_{i=1}^n \frac{|X_j|}{V_{n,\alpha}} \leq 1 \\
\frac{|X_j|}{V_{n,\alpha}} \quad &\leq \quad 1.
\end{aligned}
$$

Therefore $E(\frac{X_j}{V_{n,\alpha}})$ exists and is 0. Therefore from (2.4.5)

$$
E\Big(\frac{\sum_{i=1}^n X_i}{V_{n,\alpha}}\Big)^2 = \sum_{i=1}^n E\Big(\frac{X_i^2}{V_{n,\alpha}^2}\Big) := E(Z_n) \text{ say.} \qquad (2.4.7)
$$

Now,

$$Z_n = \sum_{i=1}^{n} \frac{X_i^2}{V_{n,\alpha}^2} \quad \leq \quad \sum_{i=1}^{n} \frac{X_i^2}{V_{n,2}^2} = 1 \text{ by Lemma 4} \tag{2.4.8}$$

Define $Y_i = sgn(X_i)|X_i|^{\frac{\alpha}{2}}$. By Lemma 2, $Y_i \in DAN$. Now

$$\begin{aligned} E(Z_n^{\alpha}) = E\Big(\frac{\sum_{i=1}^{n} X_i^2}{V_{n,\alpha}^2}\Big)^{\alpha} \quad &\leq \quad E\Big(\frac{\sum_{i=1}^{n} |X_i|^{2\alpha}}{(\sum_{i=1}^{n} |X_i|^{\alpha})^2}\Big) \text{ since } \alpha \leq 1 \\ &= \quad E\Big(\frac{\sum_{i=1}^{n} Y_i^4}{(\sum_{i=1}^{n} Y_i^2)^2}\Big) = nE\Big(\frac{Y_1^4}{(\sum_{i=1}^{n} Y_i^2)^2}\Big) \\ &= \quad o(1) \text{ by Lemma 1} \end{aligned} \tag{2.4.9}$$

Therefore $Z_n \xrightarrow{L_\alpha} 0$ and hence $Z_n \xrightarrow{P} 0$. Applying the probabilistic version of DCT (see, for example, Problem 2.37 in Athreya and Lahiri [2]) together with the bound as in (2.4.8) we have $E(Z_n) \to 0$ and hence the lemma is proved. ■

We now proceed to prove the result for $1 < p = \alpha < 2$.

Lemma 6. If $1 < p = \alpha < 2$, then $\frac{\sum X_i^2}{V_{n,\alpha}^2} \xrightarrow{P} 0$.

Proof. Define as above $Y_i = sgn(X_i)|X_i|^{\frac{\alpha}{2}}$. From Lemma 2, $Y_i \in DAN$. Therefore by [50]

$$\begin{aligned} \frac{\max_{1\leq i\leq n} |Y_i|}{(\sum_{i=1}^{n} Y_i^2)^{\frac{1}{2}}} \xrightarrow{P} 0 \quad &\Leftrightarrow \quad \frac{\max_{1\leq i\leq n} |X_i|^{\frac{\alpha}{2}}}{(\sum_{i=1}^{n} |X_i|^{\alpha})^{\frac{1}{2}}} \xrightarrow{P} 0 \\ \Leftrightarrow \Big(\frac{\max_{1\leq i\leq n} |X_i|^{\frac{\alpha}{2}}}{(\sum_{i=1}^{n} |X_i|^{\alpha})^{\frac{1}{2}}}\Big)^{\frac{4}{\alpha}} \xrightarrow{P} 0 \quad &\Leftrightarrow \quad \frac{\max_{1\leq i\leq n} X_i^2}{V_{n,\alpha}^2} \xrightarrow{P} 0 \Leftrightarrow \frac{\tilde{X}_n^*}{V_{n,\alpha}^2} \xrightarrow{P} 0 \end{aligned} \tag{2.4.10}$$

where $\tilde{X}_n^* = \max_{1\leq i\leq n} X_i^2$. Using a result of Feller [23] we have that $X_i^2 \in DA(\frac{\alpha}{2})$ where

$\frac{\alpha}{2} \in (\frac{1}{2}, 1)$ Also from Proposition 9 we have that the ratio $\tilde{X_n}^* / \sum_{i=1}^{n} X_i^2$ has a limiting distribution and hence it is tight. So given $\epsilon, \eta > 0$ obtain $K_\eta > 0$ such that $P(\sum_{i=1}^{n} X_i^2 / \tilde{X}_n^* > K_\eta) < \eta$, $\forall n$ sufficiently large. Define $\delta = \epsilon / K_\eta$.Then,

$$
\begin{aligned}
P(\frac{\sum_{i=1}^{n} X_i^2}{V_{n,\alpha}^2} > \epsilon) &\leq P(\frac{\sum_{i=1}^{n} X_i^2}{V_{n,\alpha}^2} > \epsilon, \frac{\tilde{X}_n^*}{V_{n,\alpha}^2} > \delta) + P(\frac{\sum_{i=1}^{n} X_i^2}{V_{n,\alpha}^2} > \epsilon, \frac{\tilde{X}_n^*}{V_{n,\alpha}^2} \leq \delta) \\
&\leq P(\frac{\tilde{X}_n^*}{V_{n,\alpha}^2} > \delta) + P(\frac{\sum_{i=1}^{n} X_i^2}{\tilde{X}_n^*} \frac{\tilde{X}_n^*}{V_{n,\alpha}^2} > \epsilon, \frac{\tilde{X}_n^*}{V_{n,\alpha}^2} \leq \delta) \\
&\leq P(\frac{\tilde{X}_n^*}{V_{n,\alpha}^2} > \delta) + P(\frac{\sum_{i=1}^{n} X_i^2}{\tilde{X}_n^*} > \frac{\epsilon}{\delta}) \\
&= P\Big(\frac{\tilde{X}_n^*}{V_{n,\alpha}^2} > \delta\Big) + P\Big(\frac{\sum_{i=1}^{n} X_i^2}{\tilde{X}_n^*} > K_\eta\Big). \qquad (2.4.11)
\end{aligned}
$$

By the choice of K_η the second term of (2.4.11) is less than η and from (2.4.10) the first term can be made smaller than η by choosing n sufficiently large. We therefore have

$$
P(\sum_{i=1}^{n} X_i^2 / V_{n,p}^2 > \epsilon) < 2\eta \text{ for } n \text{ sufficiently large}
$$

and hence the lemma is proved. ■

For $1 < p = \alpha < 2$ we have a slightly stronger statement assuming symmetry of X_i about 0:

Lemma 7. Let $1 < p = \alpha < 2$, and $X_i's$ are symmetric about 0 and $X_i \in DA(\alpha)$. Then $\lim_{n\to\infty} E(\frac{S_n}{V_{n,\alpha}})^2 = 0$.

Proof: From (2.4.7) we have that $E(S_n/V_{n,\alpha})^2 = E(\sum_{i=1}^{n} X_i^2/V_{n,\alpha}^2)$. By Lemma 4 we have that

$$\begin{aligned} V_{n,\alpha} &\geq V_{n,2} \text{ for } 0 < \alpha \leq 2, \\ \Rightarrow \frac{\sum_{i=1}^{n} X_i^2}{\left(\sum_{i=1}^{n} |X_i|^\alpha\right)^{\frac{2}{\alpha}}} &\leq \frac{\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} X_i^2} = 1. \end{aligned}$$

And hence by applying the Lemma 6 and by BCT we have that

$$\lim_{n\to\infty} E\Big(S_n/V_{n,\alpha}\Big)^2 = 0.$$

This proves the lemma. ■

Remark 2. For $X_i \in DA(\alpha)$ symmetric about 0, we showed in Lemma 5, 6 and 7 that $Var(S_n/V_{n,p}) \to 0$ for $0 < p = \alpha < 2$. Hence it is immediate that $Var(S_{[nt]}/V_{n,p}) \to 0$. Indeed for any fixed $0 \leq t \leq 1$,

$$Var\Big(\frac{S_{[nt]}}{V_{n,p}}\Big) = E\Big(\frac{\sum_{i=1}^{[nt]} X_i^2}{V_{n,p}^2}\Big) \leq E\Big(\frac{\sum_{i=1}^{n} X_i^2}{V_{n,p}^2}\Big) = Var\Big(\frac{S_n}{V_{n,p}}\Big) \to 0.$$

for $0 < p = \alpha < 2$. The result for k dimension can be obtained from the above result. Note that the joint distribution of $(\frac{S_{[nt_1]}}{V_{n,p}}, \frac{S_{[nt_2]}}{V_{n,p}}, \ldots, \frac{S_{[nt_k]}}{V_{n,p}})$ can be obtained from the joint distribution of $(\frac{S_{[nt_1]}}{V_{n,p}}, \frac{S_{[nt_2]}-S_{[nt_1]}}{V_{n,p}}, \frac{S_{[nt_3]}-S_{[nt_2]}}{V_{n,p}}, \ldots, \frac{S_{[nt_k]}-S_{[nt_{k-1}]}}{V_{n,p}})$ by a linear transformation. It is therefore sufficient to show that the joint distribution of the latter converges to zero. Write $S_1 = \frac{S_{[nt_1]}}{V_{n,p}}, S_2 = \frac{S_{[nt_2]}-S_{[nt_1]}}{V_{n,p}}$ and $S_k = \frac{S_{[nt_k]}-S_{[nt_{k-1}]}}{V_{n,p}}$. Now consider the variance of any linear combination of them $Var(a_1S_1 + a_2S_2 + \cdots + a_kS_k)$ where $a_i's$ are any arbitrary constants. Due to independence the cross product term vanishes and by Lemmas 5 - 7 the limiting variances are zero which implies that any linear combination tends to zero in probability. Therefore $\phi_{S_1,S_2,\ldots,S_k}(a_1, a_2, \ldots, a_k) \to 1$, where $\phi_{S_1,S_2,\ldots,S_k}$ is the joint characteristic function. Applying continuity Theorem we therefore have that the limiting joint distribution of $(S_1, S_2, \ldots, S_k)$ and hence the limiting distribution of $(S_{[nt_1]}/V_{n,p}, S_{[nt_2]}/V_{n,p}, \ldots, S_{[nt_k]}/V_{n,p})$ is degenerate at 0.

REMARK 3. For $0 < p = \alpha < 1$ Lemma 5 also holds without the symmetric assumption. Indeed, since $Y_i \sim DA(2)$, where $Y_i = \mathrm{sgn}(X_i)|X_i|^{\frac{\alpha}{2}}$, we have by [50]

$$\frac{\max\limits_{1\le i\le n}|Y_i|}{(\sum\limits_{i=1}^{n} Y_i^2)^{\frac{1}{2}}} \xrightarrow{P} 0 \quad \Leftrightarrow \quad \frac{X_n^*}{V_{n,\alpha}} \xrightarrow{P} 0, \tag{2.4.12}$$

where $X_n^* = \max\limits_{1\le i\le n}|X_i|$. For $\alpha \in (0,1)$ we have by Proposition 9 that the ratio $\frac{\sum_{i=1}^{n}|X_i|}{X_n^*}$ has a limiting distribution and is therefore tight. Therefore following the same sequence of arguments as in the proof Lemma 5 we have

$$\frac{\sum\limits_{i=1}^{n}|X_i|}{V_{n,\alpha}} \xrightarrow{P} 0 \Rightarrow \frac{S_n}{V_{n,\alpha}} \xrightarrow{P} 0.$$

Now

$$|S_n| \le V_{n,1} \le V_{n,\alpha},$$

which implies by BCT that

$$E\Big(\frac{S_n}{V_{n,\alpha}}\Big) \to 0 \text{ as } n \to \infty \text{ for } 0 < p = \alpha < 1.$$

However for $1 \le p = \alpha < 2$ some form of mean correction is needed since in the case of positive random variables $S_n = V_{n,1} \ge V_{n,\alpha} \Rightarrow \frac{S_n}{V_{n,\alpha}} \ge 1$, thus cannot converge to zero in probability. Therefore assumption of symmetric about zero has been used which is sufficient but possibly not necessary.

2.4.2.3 Case 3: $p > \alpha$

The aim of this subsection is to find the limiting joint characteristic function of the process $Y_{n,p}(t)$ at time points $0 < t_1 < t_2 < \dots t_k < 1$, $\forall k \in \mathcal{N}$. Fix $k \ge 1$. Defining $m_i = [nt_i]$ $\forall i = 1,2,\dots,k$ we first find the limiting joint characteristic function of $\mathbf{S}_1 := \Big(S_{m_1}/n^{\frac{1}{\alpha}}, (S_{m_2} - S_{m_1})/n^{\frac{1}{\alpha}}, \dots, (S_{m_k} - S_{m_{k-1}})/n^{\frac{1}{\alpha}}, V_{n,p}^{p}/n^{\frac{p}{\alpha}}\Big)$. From this applying a transformation one can obtain the limiting joint distribution of

$\mathbf{S} := \Big(\frac{S_{m_1}/n^{\frac{1}{\alpha}}}{V_{n,p}/n^{\frac{1}{\alpha}}}, \frac{S_{m_2}/n^{\frac{1}{\alpha}}}{V_{n,p}/n^{\frac{1}{\alpha}}}, \ldots, \frac{S_{m_k}/n^{\frac{1}{\alpha}}}{V_{n,p}/n^{\frac{1}{\alpha}}}\Big)$. Also since,

$$\begin{aligned}
E(|Y_{n,p}(t_1) - S_{[nt_1]}/V_{n,p}|^2) &= E((nt_1 - [nt_1])^2 |X_{[nt_1]}|^2 / V_{n,p}^2) \\
&\leq E(X_{[nt_1]+1}^2 / V_{n,p}^2) \\
&\leq E(X_{[nt_1]+1}^2 / V_{n,2}^2) \quad \forall p \leq 2 \\
&= \frac{1}{n}, \quad \text{since } [nt_1] < n,
\end{aligned}$$

the difference between the two vectors $\Big(Y_{n,p}(t_1), Y_{n,p}(t_2), \ldots, Y_{n,p}(t_k)\Big)$ and $\mathbf{S}$ are asymptotically negligible. To prove that the finite dimensional distribution of the process $Y_{n,p}(\cdot)$ exists it therefore suffices to show the existence of the limiting characteristic function of $\mathbf{S}_1$, say $\phi_{\mathbf{S}_1}(u_1, u_2, \ldots, u_k, s)$.

To find the required characteristic function we proceed along the same lines as in Logan *et al.* [40]. Note that for appropriately chosen constants a_n, $a_n S_n$ and $a_n^p V_{n,p}^p$ has the same limiting distribution as that in the case when X_i belongs to the stable distribution with index α (also see Peña *et al.* [52], pg 208). So we may and do assume that X_i's belong to stable distributions (having density $g(\cdot)$) which satisfies $x^{\alpha+1} g(x) \to r$ as $x \to \infty$ and $|x|^{\alpha+1} g(x) \to l$ as $x \to -\infty$ with $r + l > 0$ which is the property of the tail of the density of an $S(\alpha)$ distribution.

Lemma 8. $\mathbf{S}_1$ converges in distribution to a random vector whose characteristic function is given by

$$\begin{aligned}
&exp\Big(\sum_{i=1}^{k} \int [exp\{iu_i y(t_i)^{\frac{1}{\alpha}} + is|y|^p (t_i)^{\frac{p}{\alpha}})\} - 1] \frac{K(y)}{y^{\alpha+1}} dy\Big) \\
\times &\lim_{m_k, n \to \infty} E\Big(e^{il|X|^p/n^{\frac{p}{\alpha}}}\Big)^{n-m_k} \qquad (2.4.13)
\end{aligned}$$

where $K(y) = \begin{cases} r & \text{if } y > 0 \\ l & \text{if } y < 0 \end{cases}$

Proof The required characteristic function is

$$\begin{aligned}
\phi_{\mathbf{S}_1}(u_1, u_2, \ldots, u_k, s) &= E\Big(\exp\{i\frac{u_1}{n^{\frac{1}{\alpha}}}S_{m_1} + i\frac{u_2}{n^{\frac{1}{\alpha}}}(S_{m_2} - S_{m_1}) + \ldots \\
&+ i\frac{u_k}{n^{\frac{1}{\alpha}}}(S_{m_k} - S_{m_{k-1}}) + i\frac{s}{n^{\frac{p}{\alpha}}}V^p_{n,p}\}\Big) \\
&= E\Big(\exp\{i\frac{u_1}{n^{\frac{1}{\alpha}}}S_{m_1} + i\frac{u_2}{n^{\frac{1}{\alpha}}}(S_{m_2} - S_{m_1}) + \ldots \\
&+ i\frac{u_k}{n^{\frac{1}{\alpha}}}(S_{m_k} - S_{m_{k-1}}) + i\frac{s}{n^{\frac{p}{\alpha}}}(V^p_{n,p} - V^p_{m_k,p} \\
&+ V^p_{m_k,p} - V^p_{m_{k-1},p} + \ldots + V^p_{m_2,p} - V^p_{m_1,p} + V^p_{m_1,p})\}\Big) \\
&= E\Big(exp\{i[\frac{u_1}{n^{\frac{1}{\alpha}}}S_{m_1} + \frac{s}{n^{\frac{p}{\alpha}}}V^p_{m_1,p}] \\
&+ i[\frac{u_2}{n^{\frac{1}{\alpha}}}(S_{m_2} - S_{m_1}) + \frac{s}{n^{\frac{p}{\alpha}}}(V^p_{m_2,p} - V^p_{m_1,p})] \\
&+ \ldots + \frac{is}{n^{\frac{p}{\alpha}}}(V^p_{n,p} - V^p_{m_k,p})\}\Big)
\end{aligned}$$

Due to independence and identical distribution of $X's$ we have

$$E[exp\{i\frac{u_1}{n^{\frac{1}{\alpha}}}S_{m_1} + i\frac{s}{n^{\frac{p}{\alpha}}}V^p_{m_1,p}\}] = E^{m_1}[exp\{iu_1\frac{X}{n^{\frac{1}{\alpha}}} + is(\frac{|X|}{n^{\frac{1}{\alpha}}})^p\}]$$

and

$$\begin{aligned}
& E[\exp\{i\frac{u_j}{n^{\frac{1}{\alpha}}}(S_{m_j} - S_{m_{j-1}}) + i\frac{s}{n^{\frac{p}{\alpha}}}(V^p_{m_j,p} - V^p_{m_{j-1},p})\}] \\
= \; & E^{m_j - m_{j-1}}[\exp\{iu_j\frac{X}{n^{\frac{1}{\alpha}}} + is(\frac{|X|}{n^{\frac{1}{\alpha}}})^p\}], \text{ for } j = 1, 2, \ldots, k.
\end{aligned}$$

Now,

$$\begin{aligned}
& E^{m_1}[\exp\{iu\frac{X}{m_1^{\frac{1}{\alpha}}}(\frac{m_1}{n})^{\frac{1}{\alpha}} + iw(\frac{|X|}{m_1^{\frac{1}{\alpha}}})^p(\frac{m_1}{n})^{\frac{p}{\alpha}}\}] \\
= \; & [\int \exp\{iu\frac{x}{m_1^{\frac{1}{\alpha}}}(\frac{m_1}{n})^{\frac{1}{\alpha}} + iw(\frac{|x|}{m_1^{\frac{1}{\alpha}}})^p(\frac{m_1}{n})^{\frac{p}{\alpha}}\}g(x)dx]^{m_1}, \quad \text{where } g(\cdot) \text{ is the density of } X, \\
= \; & [1 + \int (exp\{iu\frac{x}{m_1^{\frac{1}{\alpha}}}(\frac{m_1}{n})^{\frac{1}{\alpha}} + iw(\frac{|x|}{m_1^{\frac{1}{\alpha}}})^p(\frac{m_1}{n})^{\frac{p}{\alpha}}\} - 1)g(x)dx]^{m_1} \\
= \; & [1 + \frac{1}{m_1}\int (exp\{iuy(\frac{m_1}{n})^{\frac{1}{\alpha}} + iw|y|^p(\frac{m_1}{n})^{\frac{p}{\alpha}}\} - 1)g(m_1^{\frac{1}{\alpha}}y) \times (m_1^{\frac{1}{\alpha}}y)^{\alpha+1}\frac{dy}{y^{\alpha+1}}]^{m_1} \\
& (\text{writing } x/m_1^{\frac{1}{\alpha}} = y).
\end{aligned}$$

Since $(exp\{iuy(\frac{m_1}{n})^{\frac{1}{\alpha}} + iw|y|^p(\frac{m_1}{n})^{\frac{p}{\alpha}}\} - 1)$ is bounded by 2 and $m_1^{\frac{1}{\alpha}+1} g(m_1^{\frac{1}{\alpha}} y)$ is integrable we apply bounded convergence theorem to get

$$\lim_{m_1, n \uparrow \infty} c_{m_1,n}(u,w) = \int [exp\{iuy(t_1)^{\frac{1}{\alpha}} + iw|y|^p(t_1)^{\frac{p}{\alpha}})\} - 1] \frac{K(y)}{y^{\alpha+1}} dy,$$

where $K(y) = \lim_{m_1 \to \infty} (m_1^{\frac{1}{\alpha}} y)^{\alpha+1} g(m^{\frac{1}{\alpha}} y)$ is given by

$$K(y) = \begin{cases} r & \text{if } y > 0 \\ l & \text{if } y < 0 \end{cases}$$

by the assumption on the tail of X and

$$c_{m_1,n}(u,w) = \int (exp\{iuy(\frac{m_1}{n})^{\frac{1}{\alpha}} + iw|y|^p(\frac{m_1}{n})^{\frac{p}{\alpha}}\} - 1) g(m^{\frac{1}{\alpha}} y) m^{\frac{1}{\alpha}+1} dy.$$

Therefore

$$\begin{aligned} & \lim_{m_1, n \to \infty, m_1/n \to t_1} E^{m_1} [exp\{iu \frac{X}{m_1^{\frac{1}{\alpha}}} (\frac{m_1}{n})^{\frac{1}{\alpha}} + iw(\frac{|X|}{m_1^{\frac{1}{\alpha}}})^p (\frac{m_1}{n})^{\frac{p}{\alpha}}\}] \\ = & \lim_{m_1, n \to \infty, m_1/n \to t_1} [1 + \frac{c_{m_1,n}}{m_1}(u,w)]^{m_1} \\ = & \ exp\{\lim_{m_1, n \to \infty, m_1/n \to t_1} c_{m_1,n}(u,w)\}. \end{aligned}$$

The same thing can be done for $E^{m_j - m_{j-1}}[\exp\{iu_j \frac{X}{n^{\frac{1}{\alpha}}} + is \frac{|X|^p}{n^{\frac{p}{\alpha}}}\}]$. Let us call it $c_{m_{j-1}, m_j, n}(u_j, s)$ for $j = 1, 2, \ldots, k$.
Therefore

$$\begin{aligned} & \lim_{m_1, m_2, \ldots, m_k, n \to \infty} \phi_{\mathbf{S}_1}(u_1, u_2, \ldots, u_k, s) \\ = & \ exp\Big(\sum_{i=1}^{k} \int [exp\{iu_i y(t_i)^{\frac{1}{\alpha}} + is|y|^p(t_i)^{\frac{p}{\alpha}})\} - 1] \frac{K(y)}{y^{\alpha+1}} dy\Big) \\ \times & \lim_{m_k, n \to \infty} E^{n-m_k}\Big(e^{il|X|^p/n^{\frac{p}{\alpha}}}\Big). \quad \blacksquare \end{aligned}$$

REMARK 4. The second limit is the limit of the characteristic function of $\frac{1}{n^{\frac{p}{\alpha}}}\sum_{i=1}^{n-m_k}|X_i|^p$ where X_i's are identical and independently distributed as a stable distribution with index α . Using the fact that $\frac{n-m_k}{n} \to 1-t_k$ and $|X|^p$ is stable with index $\frac{\alpha}{p}$, by Slutsky's Lemma we have that $\frac{1}{n^{\frac{p}{\alpha}}}\sum_{i=1}^{n-m_k}|X_i|^p = \left(\frac{n-m_k}{n}\right)^{\frac{p}{\alpha}}\frac{1}{(n-m_k)^{\frac{p}{\alpha}}}\sum_{i=1}^{n-m_k}|X_i|^p \xrightarrow{\mathcal{L}} (1-t_k)Y$, where $Y \in S(\frac{p}{\alpha})$. Hence by Levy's continuity Theorem the last limit exists and we have shown that the limiting characteristic function in the left side of Equation (2.4.13) exists for $p > \alpha$. (We have not identified the limiting distribution. For identification one can see the procedure followed in Logan *et al.* [40]).

REMARK 5. For $p = \alpha = 2$ the convergence of finite dimensional distribution of $Y_{n,p}(t)$ can be obtained from Proposition 10 since the Self-Normalized sums is converging in probability to the Wiener process properly scaled in the sup norm metric.

2.4.3 Tightness

We first state and prove a lemma that will be used in Lemma 2.

LEMMA 9. $S_k/V_{n,p}$ is a martingale with respect to the filtration

$$\mathcal{F}_{k,n} = \sigma\{\frac{X_1}{V_{n,p}}, \frac{X_2}{V_{n,p}}, \ldots, \frac{X_k}{V_{n,p}}\},\ k = 1, 2, \ldots, n$$

for every fixed $n \in \mathbb{N}$.

Proof: Let us introduce the Rademacher variables $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ i.i.d where $P(\epsilon_i = 1) = P(\epsilon_i = -1) = \frac{1}{2}$, independent of $X_i's$. Since X_i is symmetric about zero the distribution of X_i is same as $X_i^* := X_i\epsilon_i$ and the distribution of S_n is same as the distribution of

$S_n^* := \sum\limits_{i=1}^{n} X_i \epsilon_i$. Then

$$\begin{aligned} E(\frac{S_{k+1}}{V_{n,p}}|\mathcal{F}_{k,n}) &= E(\frac{S_{k+1}^*}{V_{n,p}}|\mathcal{F}_{k,n}) \\ &= E\big(E(\frac{S_k^* + X_{k+1}^*}{V_{n,p}}|\epsilon_i,\ i=1,\ldots,k)|\mathcal{F}_{k,n}\big) \\ &= E\big(\frac{S_k^*}{V_{n,p}} + E(\frac{X_{k+1}^*}{V_{n,p}}|\epsilon_i,\ i=1\ldots,k)|\mathcal{F}_{k,n}\big) \\ &= E\big(\frac{S_k}{V_{n,p}}|\mathcal{F}_{k,n}\big) = \frac{S_k}{V_{n,p}}, \end{aligned}$$

since

$$E(\frac{X_{k+1}^*}{V_{n,p}}|\epsilon_i, i=1,2,\ldots,k) = E(E(\frac{X_{k+1}^*}{V_{n,p}}|X_{k+1}, V_{n,p})|\epsilon_i, i=1,\ldots,k) = 0.$$

■

For tightness of the process $Y_{n,p}(\cdot)$ we have the following Theorem.

THEOREM 2. The process $\{Y_{n,p}(\cdot)\}$ is tight iff $p \le \alpha \le 2$.

Proof: We first prove the *if* part and then the *only if* part.
***If* part:** The process $Y_{n,p}(\cdot)$ is tight if $p \le \alpha \le 2$.

Proof: From Theorem 7.4 of Billingsley, [14] the process $Y_{n,p}(\cdot)$ is tight if:

$$P(\omega_X(\delta) \ge 3\epsilon) \quad \le \quad \sum_{i=1}^{v} P(\sup_{t_{i-1} \le s < t_i} |X(s) - X(t_{i-1})| \ge \epsilon)$$

for any partition $0 = t_0 < t_1 < t_2 < \ldots < t_v = 1$ such that $\min_{1<i<v}(t_i - t_{i-1}) \ge \delta$, and $\omega_X(\delta)$ is the modulus of continuity defined by

$$\omega_x(\delta) = \sup_{|s-t| \le \delta} |x_s - x_t|.$$

Take partition $t_i = m_i/n$ where $0 = m_0 < m_1 < \ldots < m_v = n$. By the definition of the process in (1.2.1) we have that $\sup_{t_{i-1}<s<t_i} |Y_{n,p}(s) - Y_{n,p}(t_{i-1})| = \max_{m_{i-1} \le k < m_i} \frac{|S_k - S_{m_{i-1}}|}{V_{n,p}}$.

Therefore,

$$P(\omega_{Y_{n,p}}(\delta) \geq 3\epsilon) \quad \leq \quad \sum_{i=1}^{v} P[\max_{m_{i-1} \leq k < m_i} |S_k - S_{m_{i-1}}| \geq \epsilon V_{n,p}].$$

By the i.i.d. property of the sequence $\{X_n\}$ the RHS of the above inequality is the same as:

$$\sum_{i=1}^{v} P[\max_{k < m_i - m_{i-1}} |S_k| > \epsilon V_{n,p}].$$

Choose $m_i = mi$ where m is an integer satisfying $m = \lceil n\delta \rceil$ and $v = \lceil n/m \rceil$. With this choice $v \to 1/\delta < 2/\delta$. Therefore for sufficiently large n,

$$\begin{aligned} P(\omega_{Y_{n,p}}, \delta \geq 3\epsilon) \quad &\leq \quad vP(\max_{k \leq m} |S_k|/V_{n,p} > \epsilon) \\ &\leq \quad 2/\delta \;\; P(\max_{k \leq m} |S_k|/V_{n,p} > \epsilon). \end{aligned}$$

For fixed $n \geq 1$, define a finite filtration by $\mathcal{F}_{k,n} = \sigma\{\frac{X_1}{V_{n,p}}, \frac{X_2}{V_{n,p}}, \ldots, \frac{X_k}{V_{n,p}}\}, \;\; k = 1, 2, \ldots, n$. Since $X_i's$ are symmetric about zero we have that $S_k/V_{n,p}$ is a martingale w.r.t the filtration $\mathcal{F}_{k,n}$ for $k = 1, 2 \ldots, n$ (see Lemma 9). Also since $\alpha < 2$ the ratio

$$\begin{aligned} \frac{V_{m,p}}{V_{n,p}} \quad &= \quad \left(\frac{\sum_{i=1}^{m} |X_i|^p}{\sum_{i=1}^{n} |X_i|^p}\right)^{\frac{1}{p}} \\ &= \quad \left(\frac{m}{n}\right)^{\frac{1}{p}} \left(\frac{h(m)}{h(n)}\right)^{\frac{1}{p}} \left(\frac{\frac{1}{mh(m)} \sum_{i=1}^{m} |X_i|^p}{\frac{1}{nh(n)} \sum_{i=1}^{n} |X_i|^p}\right)^{\frac{1}{p}} \\ &= \quad \left(\frac{m}{n}\right)^{\frac{1}{p}} \left(\frac{h(m)}{h(n)}\right)^{\frac{1}{p}} \left(\frac{\frac{1}{mh(m)} \sum_{i=1}^{m} Y_i^2}{\frac{1}{nh(n)} \sum_{i=1}^{n} Y_i^2}\right)^{\frac{1}{p}} \quad \text{where } Y_i = |X_i|^{\frac{p}{2}} \end{aligned}$$

$$. \tag{2.4.14}$$

Now $Y_i \in DAN$ from Lemma 2. From Proposition 8 we have that if $Y_i \in DAN$ then $\frac{\sum_{i=1}^{n} Y_i^2}{nh(n)} \xrightarrow{P} 1$. Now

$$\begin{aligned} \frac{h(m)}{h(n)} &= \frac{h(\lceil n\delta \rceil)}{h(n)} = \frac{h(n\delta - x_n)}{h(n)}, \quad \text{for some } 0 < x_n < 1, \\ &= \frac{h(n(\delta - \frac{x_n}{n}))}{h(n)} \ . \end{aligned} \tag{2.4.15}$$

For fixed δ, $\delta - \frac{x_n}{n}$ lies in some compact interval and from Proposition 6 we have that the convergence of $\frac{L(\lambda x)}{L(x)}$ to one is uniform (with respect to λ) for λ lying in any compact interval. Hence $\left(\frac{h(m)}{h(n)}\right)^{\frac{1}{p}}$ converges to 1 as $n \to \infty$. Since $\frac{\lceil n\delta \rceil}{n} \to \delta$ as $n \to \infty$, applying Slutsky's lemma we have that

$$V_{m,p}/V_{n,p} \ \xrightarrow{P} \ \delta^{\frac{1}{p}} \ .$$

Now,

$$\frac{1}{\delta} P(\max_{k \le m} |S_k|/V_{n,p} > \epsilon) \ = \ \frac{1}{\delta} P(\max_{k \le m} \frac{|S_k|}{V_{m,p}} \frac{V_{m,p}}{V_{n,p}} > \epsilon),$$

therefore writing $Z_m = \max_{k \le m} \frac{|S_k|}{V_{m,p}}$ and $Y_m = \frac{V_{m,p}}{V_{n,p}}$ we have

$$
\begin{aligned}
& \frac{1}{\delta} P(\max_{k<m} |S_k|/V_{n,p} > \epsilon) \\
= \quad & \frac{1}{\delta} P(Z_m Y_m > \epsilon) \\
= \quad & \frac{1}{\delta}\{P(Z_m Y_m > \epsilon, Y_m \le 2\delta^{\frac{1}{p}}) + P(Z_m Y_m > \epsilon, Y_m > 2\delta^{\frac{1}{p}})\} \\
\le \quad & \frac{1}{\delta}\{P(Z_m Y_m > \epsilon, Y_m \le 2\delta^{\frac{1}{p}}) + P(Y_m > 2\delta^{\frac{1}{p}})\} \\
\le \quad & \frac{1}{\delta}\{P(Z_m > \epsilon/2\delta^{\frac{1}{p}}) + P(Y_m > 2\delta^{\frac{1}{p}})\} \\
\le \quad & \frac{1}{\delta}\{P(Z_m > \epsilon/2\delta^{\frac{1}{p}}) + \eta\} \quad \text{(choosing sufficiently large m} \\
& \text{such that } P(Y_m - \delta^{\frac{1}{p}} > \delta^{\frac{1}{p}}) < \eta) \\
\le \quad & \frac{1}{\delta}\{(4\delta^{\frac{2}{p}}/\epsilon^2) E(S_m/V_{m,p})^2 + \eta\}, \text{ by Doob's inequality,} \\
& \text{for non-negative submartingales} \\
= \quad & (4\delta^{\gamma}/\epsilon^2) E(S_m/V_{m,p})^2 + \eta/\delta, \qquad (2.4.16)
\end{aligned}
$$

where $\gamma = \frac{2}{p} - 1$. Now, for $p \le \alpha < 2$, or for , $p < \alpha = 2$, $E(S_m/V_{m,p})^2$ tends to zero (see Section 2.4.2.1, 2.4.2.2). Taking $m \to \infty$, (since $m = \lceil n\delta \rceil$) the right hand side in (2.4.16) can be made arbitrarily small. This proves the lemma for $p \le \alpha < 2$ and $p < \alpha = 2$. For the case $p = \alpha = 2$, the lemma holds by Giné *et al.* [27] since it has been shown that the Self-Normalized sums converges to the Normal distribution for $p = \alpha = 2$. ■.

Before proving the only if part we prove the following lemma.

LEMMA 10. $\{Y_{n,p}(\cdot)\}$ is tight $\Rightarrow \max_{1 \le i \le n} \frac{|X_i|}{V_{n,p}} \xrightarrow{P} 0$.

Proof. From Theorem 7.3 of Billingsley [14], the process $Y_{n,p}(\cdot)$ is tight is equivalent to $\forall \epsilon > 0,\ \forall \eta > 0,\ \exists n_0$ and $0 < \delta < 1$ such that

$$
P\Big(\sup_{|t-s|<\delta} |Y_{n,p}(s) - Y_{n,p}(t)| \ge \epsilon\Big) \quad \le \quad \eta,\ \forall n \ge n_0. \qquad (2.4.17)
$$

Assume that the hypothesis is true, which means that for every $\epsilon, \eta > 0, \exists 0 < \delta < 1$ such that (2.4.17) holds. Choose n_0 sufficiently large so that $\frac{1}{n} < \delta \ \ \forall n > n_0$. Then we have

$$P(\sup_{|t-s|<\frac{1}{n}} |Y_{n,p}(t) - Y_{n,p}(s)| > \epsilon) \ < \ P(\sup_{|t-s|<\delta} |Y_{n,p}(t) - Y_{n,p}(s)| > \epsilon) \leq \eta \tag{2.4.18}$$

Now by definition of the process $Y_{n,p}(\cdot)$,

$$\begin{aligned}
\sup_{|t-s|<\frac{1}{n}} |Y_{n,p}(t) - Y_{n,p}(s)| &\geq \max_{1\leq i\leq n} \frac{|X_i|}{V_{n,p}} \\
\Rightarrow P(\max_{1\leq i\leq n} \frac{|X_i|}{V_{n,p}} > \epsilon) &< P(\sup_{|t-s|<\delta} |Y_{n,p}(t) - Y_{n,p}(s)| > \epsilon) \\
\Rightarrow P(\max_{1\leq i\leq n} \frac{|X_i|}{V_{n,p}} > \epsilon) &< \eta, \quad \forall n > n_0,
\end{aligned}$$

by Equation (2.4.18). ■

Remark 6. The converse is not necessarily true. To see this assume that $\max_{1\leq i\leq n} \frac{|X_i|}{V_{n,p}} \xrightarrow{P} 0$. Assume that there exists a δ_1 such that (2.4.17) holds. Given such a $\delta_1 > 0$, for any integer m we can get an n such that $\frac{m}{n} < \delta_1$. Then for such a m, n we have $|Y_{n,p}(t) - Y_{n,p}(s)| \leq (\max_{1\leq i\leq n} \sum_{j=1}^{m} |X_{i+j}|)/(V_{n,p})$. But the hypothesis does not guarantee that the right hand side converges to zero in probability.

We use the above lemma to prove the necessary part in the following lemma.

***Only if* part** For $2 \geq p > \alpha$ the process is not tight.

Proof:

For $2 \geq p > \alpha$ observe that

$$\max_{1\leq i\leq n} \frac{|X_i|}{V_{n,p}} \xrightarrow{P} 0 \Leftrightarrow \left(\max_{1\leq i\leq n} \frac{|X_i|}{V_{n,p}}\right)^p \xrightarrow{P} 0 \Leftrightarrow \max_{1\leq i\leq n} \frac{|X_i|^p}{\sum |X_i|^p} \xrightarrow{P} 0.$$

But $|X_i|^p \in DA(\gamma)$, where $\gamma = \frac{\alpha}{p} < 1$, for which Darling [19], Theorem 5.1, says that if $Y_i \in DA(\gamma)$ where $\gamma < 1$ then $\max_{1\leq 1\leq n} \frac{|Y_i|}{\sum |Y_i|}$ converges in distribution to a non-degenerate random variable G whose characteristic function is identified in the same

paper. Thus, $\max_{1\le i\le n}\frac{|X_i|^p}{\sum|X_i|^p}$ does not go to zero in probability. Hence, $max_{1\le i\le n}\frac{X_i}{V_{n,p}}$ cannot converge to zero in probability and therefore from Lemma 10 the process cannot be tight.
For $p > 2 = \sup\{\, p_1 : E|X_1|^{p_1} < \infty\}$ then $|X_i|^p \in DA(\frac{2}{p})$ and the proof is exactly similar. Combining the *if* and the *only if* part we have the proof of Lemma 2. ■

REMARK 7. Consider the case when $p > p_1 > \alpha = 2$ where p_1 is such that $E(|X|^{p_1}) < \infty$. In that case

$$\begin{aligned} |Y_n| &= \frac{|S_n|}{V_{n,p}} \ge \frac{|S_n|}{V_{n,p_1}} \\ &= \left(n^{\frac{1}{2}-\frac{1}{p_1}}\right)\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n} X_i\right|/\left(\frac{1}{n}\sum_{i=1}^{n}|X_i|^{p_1}\right)^{\frac{1}{p_1}} \\ &\to \infty, \end{aligned}$$

which proves that the f.d.d's of Y_n does not exists.

2.5 The stochastic process of Basak and Dasgupta [4]

As an example let us consider another way of normalizing the SNS which is due to Basak and Dasgupta [4]. Let S_j be defined as earlier, with $X_i \sim DAN$. Let $V_{j,2}^2 = \sum_{i=1}^{j} X_i^2$ and $Y_j = S_j/V_{j,2}$. Then define the stochastic process $Y^n(\cdot)$ as

$$\begin{aligned} Y^n(0) &= Y_n \\ Y^n(\sum_{j=n+1}^{l} b_j^2) &= Y_l,\ l \ge n+1, \end{aligned} \tag{2.5.19}$$

where $b_{j+1}^2 = \frac{1}{j+1}$ $(= E(\frac{X_{j+1}^2}{V_{j+1}^2})$) since $E\left(\frac{X_i^2}{V_{j+1,2}^2}\right)$ is same for all $i = 1, 2, \ldots, j+1$ and $\sum_{i=1}^{j+1}\frac{X_i^2}{V_{j+1,2}^2} = 1$. At intermediate points the process is obtained by joining the nearest points linearly. Since $\max\{k : \sum_{j=n}^{k}\frac{1}{j+1} \le t\} \sim [ne^t]$, for large n, the limiting distribution of $Y^n(t)$ is same as the limiting distribution of

$$Y_{[ne^t]} = \frac{S_{[ne^t]}}{V_{[ne^t]}}.$$

With this normalization the authors had the following result :

PROPOSITION 12. Let $\{X_i\}_{i\geq 1}$ be an i.i.d sample from the domain of attraction of a Normal distribution. Let $Y^n(\cdot)$ be defined as earlier. Then $Y^n(\cdot)$ converges weakly in $C[0,1]$ to the stationary Ornstein Uhlenbeck process with covariance function $e^{-\frac{1}{2}|t-s|}$.

REMARK 8. In Equation (2.5.19) the random variable $Y_1, Y_2, \ldots, Y_n$ used in defining the value of the first n stochastic processes $\{Y^1(t), Y^2(t), \ldots, Y^n(t)\}$ at $t = 0$ is no longer used in defining the subsequent stochastic processes with index $n+1$ and upwards. Also the way the stochastic processes $Y_n(\cdot)$ are defined they are stationary. Indeed, the distribution of $Y^n(t)$ is approximately same as $S_{[ne^t]}/V_{[ne^t]}$ for large t. From Proposition 8 the joint distribution of

$$\Big(\frac{S_{[ne^{t_1}]}}{V_{[ne^{t_1}]}}, \frac{S_{[ne^{t_2}]} - S_{[ne^{t_2}]}}{\sqrt{V^2_{[ne^{t_2}]} - V^2_{[ne^{t_1}]}}}, \ldots, \frac{S_{[ne^{t_k}]} - S_{[ne^{t_{k-1}}]}}{\sqrt{V^2_{[ne^{t_k}]} - V^2_{[ne^{t_{k-1}}]}}}\Big)$$

converges to the k-fold product of independent $N(0,1)$. Also note that

$$\Big(\frac{V_{[ne^{t_1}]}}{V_{[ne^{t_i}]}}, \frac{\sqrt{V^2_{[ne^{t_2}]} - V^2_{[ne^{t_1}]}}}{V_{[ne^{t_i}]}}, \ldots, \frac{\sqrt{V^2_{[ne^{t_i}]} - V^2_{[ne^{t_{i-1}}]}}}{V_{[ne^{t_i}]}}\Big)$$

jointly converges in probability to

$$\Big(\frac{e^{\frac{t_1}{2}}}{e^{\frac{t_1}{2}}}, \frac{\sqrt{e^{t_2} - e^{t_1}}}{e^{\frac{t_i}{2}}}, \ldots \frac{\sqrt{e^{t_i} - e^{t_{i-1}}}}{e^{\frac{t_i}{2}}}\Big).$$

Hence writing,

$$\frac{S_{[ne^{t_i}]}}{V_{[ne^{t_i}]}} = \frac{V_{[ne^{t_1}]}}{V_{[en^{t_i}]}} \frac{S_{[ne^{t_1}]}}{V_{[ne^{t_1}]}} + \sum_{l=2}^{i} \frac{\sqrt{V^2_{[ne^{t_l}]} - V^2_{[ne^{t_{l-1}}]}}}{V_{[ne^{t_i}]}} \frac{S_{[ne^{t_l}]} - S_{[ne^{t}_{l-1}]}}{\sqrt{V^2_{[ne^{t_l}]} - V^2_{[ne^{t_{l-1}}]}}}$$

and using Slutsky's Theorem, the limiting distribution of

$$\left(\frac{S_{[ne^{t_1}]}}{V_{[ne^{t_1}]}}, \ldots, \frac{S_{[ne^{t_k}]}}{V_{[ne^{t_k}]}}\right)$$

is seen to be multivariate Normal with the covariance between $\frac{S_{[ne^{t_i}]}}{V_{[ne^{t_i}]}}$ and $\frac{S_{[ne^{t_j}]}}{V_{[ne^{t_j}]}}$ with $i < j$ being:

$$\frac{e^{t_1}}{e^{t_i/2}e^{t_j/2}} + \sum_{m=2}^{i} \frac{e^{t_m} - e^{t_{m-1}}}{e^{t_i/2}e^{t_j/2}} = e^{-\frac{1}{2}(t_j - t_i)}.$$

In the above the random variables came from DAN. A natural question is whether a converse to their ([6]) result exist. To do so, we define a new stochastic process $\tilde{Y}^n(\cdot)$ where the random variables X_i come from $DA(\alpha)$ and the process is defined as:

$$\tilde{Y}^n_{\alpha,p}(t) = \frac{S_{[ne^t]}}{V_{[ne^t],p}},$$

where $V^p_{[ne^t],p} = \sum_{i=1}^{[ne^t]} |X_i|^p$ and $X_i \sim DA(\alpha)$ are symmetric about 0 and $\alpha \in (0, 2]$ and $p > 0$. Now since both $S_{[ne^t]}$ and $V_{[ne^t]}$ has the same number of terms in the summation one can argue as in Section 2.4.2 that the process will not have a non-trivial limiting distribution unless $p > \alpha$. Arguing along the same lines as in Section 2.4.3 that the process is not tight unless $p \leq \alpha$. Combining these two we get that the only case when the process has a non-trivial limiting distribution is when $p = \alpha = 2$. The limiting distribution of the process is an Ornstein Uhlenbeck process as stated in Proposition 12.

2.6 Summary

In this chapter we looked at the functional form of the classical Self-Normalized sums $S_n/V_{n,p}$. This quantity arises naturally when defining the t statistics. It is known from the works of Efron [20], Giné *et al.* [27], Csörgő *et al.* [17] that the natural class of distributions for which we can expect a non-trivial limiting distribution for the SNS should be the domain of attraction of the stable distribution. The above papers typically dealt with L_2 normalization where the random variables also belonged to the domain of attraction of the $S(2)$ distribution.

In this chapter we have extended the above works by exploring into what happens if the random variables are not restricted to the DA(2) family. Also we looked into the possibility if we extend the scaling coefficient from 2 to p where p is any positive real. Our results proved that we cannot expect any non-trivial limiting results for the SNP if p and α are both not equal to 2.

Chapter 3

Diffusion approximation of Adaptive MCMC

3.1 Introduction to AMCMC

The second example of recursive system that we consider is the Adaptive Markov Chain Monte Carlo (AMCMC) algorithms. An AMCMC is a discrete time stochastic process $X_n, n = 1, 2, \ldots$ on a general state space, say $\mathcal{X}$, whose transition kernel is not only dependent on the present state but also on the infinite history (or it might be dependent on a parameter that is a function of the previous history of the chain). For example, if X_n is a AMCMC, then the transition kernel $P(X_n \in A|X_{n-1} = x)$ will depend on x and an adaptation parameter, say γ_n, taking values in some adaptation index set, say Γ. γ_n might be a function of $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$, or is a constant that change with each iteration.

AMCMC typically arise in statistical simulation using the MCMC technique. MCMC is a general strategy for generating samples $\{X_i, i = 0, 1, 2, \ldots\}$ from complex high dimensional distributions, say ψ, defined on a space $\mathcal{X} \subset \mathbb{R}^n$(assumed for simplicity to have a density with respect to the Lebesgue measure, also denoted by $\psi(\cdot)$) from which integrals of the type

$$I(f) := \int_{\mathcal{X}} f(x)\psi(x)dx$$

for some ψ- integrable function $f : \mathcal{X} \to \mathbb{R}^n$ can be approximated using the estimator

$$\hat{I}_N(f) := \frac{1}{N} \sum_{i=n_0+1}^{n_0+N} f(X_i),$$

where n_0 is the initial burn-in period and the random variable X_i are generated from a MC which converges to $\psi(\cdot)$ in the total variation norm. The main building block of this class of algorithms is the MH (Metropolis Hastings) algorithms. For the construction of such algorithms and their properties see Section 1.3.

As discussed in the introduction the simplicity of the algorithm is both its strength and weakness. The choice of the scale parameters are crucial, and bad choices could lead to the samples which do not converge fast to the required distribution.

AMCMC eschews this problem by tuning the parameters in an optimal way. That is to say that the parameters are so chosen that the convergence to $\psi(\cdot)$ becomes fast. Usually the choice of the scaling parameter is the scale parameter of the density of the transition kernel. However since the transition kernel is not the same at each iteration any arbitrarily constructed AMCMC cannot guarantee convergence. Sufficient conditions for ergodicity is given in Proposition 3 in Chapter 1.

Remark 9. In essence, condition 1 of Proposition 3 of Chapter 1 says that for any fixed $\gamma \in \Gamma$ and starting point $x \in \mathcal{X}$, the transition kernel $P_\gamma(\cdot, \cdot)$ is ergodic (see definition in Page 12). In addition, the rate of convergence to the invariant distribution is uniform over all $x \in \mathcal{X}$ and $\gamma \in \Gamma$. Condition 2 there says that the change in the transition kernel (as measured in total variation norm) over each iteration decreases to zero as $n \to \infty$ uniformly over all $x \in \mathcal{X}$.

From the definition of AMCMC (see Section 1.3 of Chapter 1) it is clear that this is an example of a recursive system. In this chapter and the next we obtain the invariant distribution of a suitably defined AMCMC, after performing the diffusion approximation procedure to the process, see, for example [48]. Our choice of the AMCMC arises from the fact that the adaptation parameter (also called *tuning/ scaling parameter*) should depend on whether the sample generated from the proposal distribution is accepted or not. If accepted, then the scaling parameter should increase by some amount and if not, the scaling parameter should decrease. The next section gives the definition of the proposed AMCMC (a partial variation of this algorithm was suggested by Prof. P. Green in a personal communication.)

3.2 Definition of the Adaptive MCMC algorithm

We assume that the target distribution $\psi(\cdot)$ is univariate and $\frac{\psi'(x)}{\psi(x)}$ grows linearly in x. (The reason for this choice is explained in Remark 13). We recall the algorithm (Algorithm 2) that was described in Section 1.3.3.3 in Chapter 1.

Algorithm 2:

1. Select arbitrary $\{X_0, \theta_0, \xi_0\} \in \mathbb{R} \times (0, \infty) \times \{0, 1\}$ where $\mathbb{R}$ is the state space which may be the real line or an interval of the same. Set $n = 1$.
2. Propose a new move, say Y, where $\quad Y \sim N(X_{n-1}, \theta_{n-1})$.
3. Accept the new point with probability $\alpha(X_{n-1}, Y) = \min\{1, \frac{\psi(Y)}{\psi(X_{n-1})}\}$.
   If the point is accepted, set $X_n = Y,\ \xi_n = 1$; else $X_n = X_{n-1},\ \xi_n = 0$.
4. $\theta_n = \theta_{n-1} e^{\frac{1}{\sqrt{n}}(\xi_n - q)}, \quad q > 0, \quad \Leftrightarrow \log(\theta_n) = \log(\theta_{n-1}) + \frac{1}{\sqrt{n}}(\xi_n - q), \quad q > 0.$
5. $n \leftarrow n + 1$, and go to step 2.

The above algorithm is equivalent to the following:
Algorithm 2′:

1. Select arbitrary $\{X_0, \theta_0, \xi_0\} \in \mathbb{R} \times (0, \infty) \times \{0, 1\}$, where $\mathbb{R}$ is the state space. Set $n = 1$.
2. Given $X_{n-1}, \theta_{n-1}, \epsilon_{n-1}$ generate

$$\xi_n \sim Bernoulli\Big(\min\Big(1, \frac{\psi(X_{n-1} + \theta_{n-1}\epsilon_{n-1})}{\psi(X_{n-1})}\Big)\Big)$$

   and then

$$X_n = X_{n-1} + \theta_{n-1}\xi_n\epsilon_{n-1}$$

   where $\epsilon_{n-1} \sim N(0, 1)$,

3. $\theta_n = \theta_{n-1} e^{\frac{1}{\sqrt{n}}(\xi_n - q)}, \quad q > 0, \quad \Leftrightarrow \log(\theta_n) = \log(\theta_{n-1}) + \frac{1}{\sqrt{n}}(\xi_n - q), \quad q > 0.$

4. $n \leftarrow n+1$ and go to step 2.

Let us describe the algorithm. θ_n is the proposal scaling (tuning) parameter which is adaptively tuned depending on whether the previous sample was accepted or rejected. If the sample was accepted then the proposal variance will increase allowing the chain to explore more regions in the state space. If the past sample was rejected then the variance will decrease making the move a more conservative one. Here q is a benchmark; for multivariate Normal target density, where the components are independent, the value 0.238 is often appropriate, see Gelman *et al.* [25]. For a further generalization see Bedard [9]. The tuning parameter can also be made to be dependent not only on whether the previous sample was accepted but also on the proportions of samples accepted in the history of the chain. However, this is not done in this thesis.
This algorithm is similar, in principle, to the Stochastic approximation (SA) procedure, see Monro and Robbins [47], for an introduction to the SA procedure.
For the rest of this chapter we try to prove the ergodicity of the proposed AMCMC algorithm.

Remark 10.
1. Note that the Diminishing Condition in Proposition 3 is satisfied since from the definition of the algorithm $\log(\theta_{n+1}) - \log(\theta_n) = o_p(\frac{1}{\sqrt{n}})$. Therefore $||P_{\theta_{n+1}}(\mathbf{x}, \cdot) - P_{\theta_n}(\mathbf{x}, \cdot)||_{TV} \to 0$.

2. It is difficult to verify that the rate of convergence is uniform over all choices of (x, γ). Therefore the result in Proposition 3 cannot be applied directly in that form.

Our aim in this chapter is to embed the discrete time chain into a continuous time stochastic process. This method, called the diffusion approximation, is quite common in probability theory. It has been applied to diverse fields, for example, econometric modelling (Nelson [48]), branching processes (Ethier and Kurtz [22]), etc. Also the advantage is that we can apply standard tools in continuous time stochastic processes, which are not available for discrete time AMCMC. Diffusion approximation was also applied by Gelman *et al.* in their paper [25], where they tried to obtain the optimal scaling proposal of the (standard, multidimensional) MH algorithm. The next section gives details of the technique.

3.3 Diffusion Approximation

In this section we first present conditions developed by Stroock and Varadhan [65] for a sequence of stochastic processes satisfying a stochastic difference equations to converge weakly to an Itô Process.

Here is the formal set up: Let $D([0,\infty),\mathbb{R}^n)$ be the space of mappings from $[0,\infty)$ into $\mathbb{R}^n$ that are continuous from the right with left limits and let $\mathcal{B}(\mathbb{R}^n)$ denote Borel sets in $\mathbb{R}^n$. D is a metric space when endowed with the Skorokhod metric (see Billingsley [14]). For each $h > 0$ let $\mathcal{M}_{kh}$ be the σ-algebra generated by the random variables $\mathbf{X}_{0,h}, \mathbf{X}_{h,h}, \ldots, \mathbf{X}_{kh,h}$ for $k \geq 1$ and let ν_h be a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. For each $h > 0$ and each $k \geq 1$ let $\Pi_{kh,h}$ be a transition kernel for a homogeneous Markov chain i.e.,

1. $\Pi_{kh,h}(\mathbf{x},\cdot)$ is a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ for all $\mathbf{x} \in \mathbb{R}^n$;
2. $\Pi_{kh,h}(\cdot, A)$ is a $\mathcal{B}(\mathbb{R}^n)$ measurable for all $A \in \mathcal{B}(\mathbb{R}^n)$.

For each $h > 0$, let P_h be the probability measure on $D([0,\infty),\mathbb{R}^n)$ such that

$$P_h\Big(\mathbf{X}_{0,h} \in A\Big) = \nu_h(A) \;\; \forall\; A \in \mathcal{B}(\mathbb{R}^n), \tag{3.3.1}$$

$$P_h\Big(\mathbf{X}_{t,h} = \mathbf{X}_{kh,h},\;\; kh \leq t \leq (k+1)h\Big) = 1, \tag{3.3.2}$$

$$P_h\Big(\mathbf{X}_{(k+1)h,h} \in A|\mathcal{M}_{kh}\Big) = \Pi_{kh,h}(\mathbf{X}_{kh,h}, A). \tag{3.3.3}$$

almost surely under P_h $\forall\; k \geq 0$ and $A \in \mathcal{B}(\mathbb{R}^n)$.

For each $h > 0$, equation (3.3.1) specifies the distribution of the random starting point. In Equation (3.3.2) we construct a continuous time process from the discrete time process by making $X_{t,h}$ a step function with jumps at time $h, 2h, 3h, \ldots$ etc. Equation (3.3.3) states that for a fixed $h > 0$, $\{X_{kh,h}, k \geq 1\}$ is a Markov Chain with $\Pi_{kh,h}(\cdot,\cdot)$ as the transition kernel.

We next define the infinitesimal diffusion and drift coefficients for any $t, h > 0$ as :

$$\begin{aligned}
a_h(\mathbf{x}, t) &= h^{-1} \int_{\mathbb{R}^n} (\mathbf{y}-\mathbf{x})(\mathbf{y}-\mathbf{x})' \Pi_{[t/h]h,h}(\mathbf{x}, d\mathbf{y}) \\
&= h^{-1} D(\mathbf{X}_{(k+1)h,h} | \mathbf{X}_{kh,h} = \mathbf{x}) \text{ for any } k \geq 1; \\
b_h(\mathbf{x}, t) &= h^{-1} \int_{\mathbb{R}^n} (\mathbf{y}-\mathbf{x}) \Pi_{[t/h]h,h}(\mathbf{x}, d\mathbf{y}) = h^{-1} E(\mathbf{X}_{(k+1)h,h} - \mathbf{x} | \mathbf{X}_{kh,h} = \mathbf{x}) \text{ for any } k \geq 1; \\
\Delta_{h,\epsilon}(\mathbf{x}, t) &= h^{-1} \int_{||\mathbf{y}-\mathbf{x}|| > \epsilon} \Pi_{[t/h]h,h}(\mathbf{x}, d\mathbf{y}) \\
&= h^{-1} P(||\mathbf{X}_{(k+1)h,h} - \mathbf{X}_{kh,h}|| > \epsilon \mid \mathbf{X}_{kh,h} = \mathbf{x}) \text{ for any } k \geq 1, \qquad (3.3.4)
\end{aligned}$$

where $D(\mathbf{X}_{(k+1)h,h} | \mathbf{X}_{kh,h} = \mathbf{x})$ and $E(\mathbf{X}_{(k+1)h,h} - \mathbf{x} | \mathbf{X}_{kh,h} = \mathbf{x})$ are the conditional dispersion and conditional expected deviation given that the value of $\mathbf{X}_{kh,h}$ is $\mathbf{x}$ respectively. $a_h(\mathbf{x}, t)$ and $b_h(\mathbf{x}, t)$ are measures of the second moment and drift per unit of time respectively. $\Delta_{h,\epsilon}(\mathbf{x}, t)$ is the conditional probability of a jump of size ϵ or greater per unit of time. The convergence results that we present below will require that $a_h(\mathbf{x}, t)$ and $b_h(\mathbf{x}, t)$ converge to a finite limits and $\Delta_{h,\epsilon}(\mathbf{x}, t)$ goes to zero for all $\epsilon > 0$ as $h \downarrow 0$. In particular we assume the following, see [65]:

Assumptions

1. There exists a locally bounded measurable mapping $a(\mathbf{x}, t) : \mathbb{R}^n \times [0, \infty) \to M^+_{n \times n}$ which are continuous in x for each $t \geq 0$, and $b(\mathbf{x}, t) : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}^n$ such that:

$$\begin{aligned}
\lim_{h \downarrow 0} ||a_h(\mathbf{x}, t) - a(\mathbf{x}, t)|| &= 0; \\
\lim_{h \downarrow 0} ||b_h(\mathbf{x}, t) - b(\mathbf{x}, t)|| &= 0; \\
\lim_{h \downarrow 0} \Delta_{h,\epsilon}(\mathbf{x}, t) &= 0,
\end{aligned}$$

   where $M^+_{n \times n}$ denotes the space of all $n \times n$ non-negative definite matrices and $|| \cdot ||$ is the matrix/vector norm defined as:

$$||A|| = \begin{cases} [A^T A]^{\frac{1}{2}} & \text{if } A \text{ is a column vector} \\ [trace(A^T A)]^{\frac{1}{2}} & \text{if } A \text{ is a matrix.} \end{cases}$$

2. There exists a locally bounded measurable mapping $\sigma(\mathbf{x}, t)$ form $\mathbb{R}^n \times [0, \infty) \to M_{n \times n}$ which are continuous in x for each $t \geq 0$, such that for all $\mathbf{x} \in \mathbb{R}^n$ and all

$t \geq 0$,

$$a(\mathbf{x}, t) = \sigma(\mathbf{x}, t)\sigma(\mathbf{x}, t)',$$

where $M_{n \times n}$ denotes the space of all $n \times n$ matrix.

3. As $h \downarrow 0$, $X_{0,h}$ converges in distribution to a random variable X_0 with a probability measure ν_0 on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$;

4. $\nu_0, a(\mathbf{x}, t)$ and $b(\mathbf{x}, t)$ uniquely specify the distribution of a diffusion process $\mathbf{X}_t$, with the initial distribution ν_0, the the diffusion matrix $a(\mathbf{x}, t)$ and the drift vector $b(\mathbf{x}, t)$.

Under the assumption we have the following Proposition. For a proof see Stroock and Varadhan [65].

Proposition 13. Under Assumptions 1 - 4, the sequence of $\mathbf{X}_{h,t}$ process defined by Equations (3.3.1) - (3.3.3) converges weakly (i.e., in distribution) as $h \downarrow 0$ to the $\mathbf{X}_t$ process defined by the stochastic integral equation

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t b(\mathbf{X}_s, s)ds + \int_0^t \sigma(\mathbf{X}_s, s)dW_{n,s} \tag{3.3.5}$$

where $W_{n,t}$ is an n-dimensional standard Brownian motion, independent of $\mathbf{X}_0$ and where for any $A \in \mathcal{B}(\mathbb{R}^n), P(X_0 \in A) = \nu_0(A)$. Such an $\mathbf{X}_t$ process exists and is unique upto a distribution.

Next we embed the discrete time Algorithm $2'$ defined in Section 3.2 in a continuous time process that has decreasing step sizes. For fixed $n \geq 1$, we partition the half line $[0, \infty)$ into sub intervals of length $\frac{1}{n}$. We start with the fixed point x_0. Now given the value of the process at time $\frac{i}{n}$, i.e., $X_n\left(\frac{i}{n}\right) = x$, we propose a value following the $N\left(x, \frac{1}{\sqrt{n}}\theta_n\left(\frac{i}{n}\right)\right)$ distribution. We have the correction factor $\frac{1}{n}$ multiplied with the variance to incorporate the diminishing adaptation condition, so that the difference between the proposal kernel at times $\frac{i}{n}$ and $\frac{i+1}{n}$ goes to zero as $n \to \infty$. This proposed value is accepted with the usual MH acceptance probability given in (1.3.2) at time $\frac{i+1}{n}$. The indicator variable denoting whether the proposed value is accepted is denoted by $\xi_n\left(\frac{i}{n}\right)$. Similar approximation is done with the tuning parameter $\theta_n(\cdot)$ starting with the initial value θ_0.

3.3.1 Embedding in continuous time of discrete AMCMC

The following gives the embedding of the discrete AMCMC into continuous times state variable $X_n(\cdot)$

$$
\begin{aligned}
X_n(0) &= x_0 \in \mathbb{R}; \\
X_n\Big(\frac{i+1}{n}\Big) &= X_n\Big(\frac{i}{n}\Big) + \frac{1}{\sqrt{n}}\theta_n\Big(\frac{i}{n}\Big)\xi_n\Big(\frac{i+1}{n}\Big)\epsilon_n\Big(\frac{i+1}{n}\Big), \quad \text{i=0, 1, } \ldots, \\
X_n(t) &= X_n\Big(\frac{i}{n}\Big), \quad \text{if } \tfrac{i}{n} \le t < \tfrac{i+1}{n} \text{ for some integer } i. \qquad (3.3.6)
\end{aligned}
$$

Here, $\xi_n(\frac{i+1}{n})$ conditionally follows the Bernoulli distribution given by:

$$
\begin{aligned}
&P\Big(\xi_n(\frac{i+1}{n}) = 1 | X_n(\frac{i}{n}),\ \theta_n\Big(\frac{i}{n}\Big),\ \epsilon_n\Big(\frac{i+1}{n}\ \Big)\Big) \\
&= \min\Big\{\frac{\psi(X_n\Big(\frac{i}{n}\Big) + \frac{1}{\sqrt{n}}\theta_n\Big(\frac{i}{n}\Big)\epsilon_n\Big(\frac{i+1}{n}\Big)\Big)}{\psi\Big(X_n\Big(\frac{i}{n}\Big)\Big)}, 1\Big\}
\end{aligned}
$$

and $\{\epsilon_n(\frac{i}{n}), i \ge 1\ \}$ are all independent $N(0,1)$ random variables. This distribution of $\xi_n(\cdot)$ comes directly from the form of the MH acceptance probability given in (1.3.2).

Tuning parameter $\theta_n(\cdot)$
The n^{th} approximation to the tuning parameter $\theta(\cdot)$ is defined as :

$$
\begin{aligned}
\theta_n(0) &= \theta_0 \in \mathbb{R}^+, \\
\theta_n\left(\frac{i+1}{n}\right) &= \theta_n\left(\frac{i}{n}\right) e^{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))}, \quad \text{i=0, 1, } \ldots, \\
\text{and } \theta_n(t) &= \theta_n(\frac{i}{n}), \quad \text{if } \tfrac{i}{n} \le t < \tfrac{i+1}{n} \text{ for some integer } i. \qquad (3.3.7)
\end{aligned}
$$

In the original discrete AMCMC the benchmark value of q, given in Step 3 of Algorithm $2'$, was kept fixed. However if that is also done in the continuous AMCMC in Equation (3.3.7) then the tuning parameter θ_n will converge to

$$
\begin{cases} \infty & \text{if } q < 1; \\ 0 & \text{if } q = 1. \end{cases}
$$

It is exactly for this reason the constants in the tuning parameter given in Equation (3.3.7) is an increasing function of n (also depending on a constant $q > 0$) that converges to 1 as $n \to \infty$. In particular, for our example, we have $q_n\left(\frac{i}{n}\right) = 1 - \frac{q}{\sqrt{n}}$ for some $q > 0$.

For comparison purposes we also embed the discrete time standard MCMC (SMCMC) in continuous times. The SMCMC algorithm is almost similar to the AMCMC, except for the fact that the tuning parameter given by $\theta(\frac{i}{n})$ corresponding to SMCMC is kept fixed at a constant level θ_0, that is unchanged in the iterations. This is done in the next subsection .

3.3.2 Embedding in continuous times of SMCMC

The continuous time process corresponding to SMCMC will therefore be :

$$\begin{aligned} X_n(0) &= x_0 \in \mathbb{R}; \\ X_n\left(\frac{i+1}{n}\right) &= X_n\left(\frac{i}{n}\right) + \frac{1}{\sqrt{n}}\theta_0\xi_n\left(\frac{i+1}{n}\right)\epsilon_n\left(\frac{i+1}{n}\right), \quad \text{i=0, 1, } \ldots, \ \theta_0 \in \mathbb{R}^+ = (0, \infty), \\ X_n(t) &= X_n\left(\frac{i}{n}\right), \quad \text{if } \tfrac{i}{n} \le t < \tfrac{i+1}{n} \ \text{ for some integer } i. \end{aligned} \tag{3.3.8}$$

where $\xi_n\left(\frac{i}{n}\right)$ has the same conditional distribution with θ_n replaced by θ_0 where θ_0 is the fixed constant that is not updated in the iterations.
The following main Theorem of this chapter tells the outcome of the diffusion approximation of the Discrete AMCMC defined through Equations (3.3.6) to (3.3.7) and that of the SMCMC defined through (3.3.8).

3.4 Main Theorem

THEOREM 3. 1. $\mathbf{Y}_n(t) := \Big(X_n(t), \ \ \theta_n(t)\Big)$ (where $X_n(t)$ and $\theta_n(t)$ is given by (3.3.6) and (3.3.7) respectively) converges weakly to a diffusion process which is the solution to the SDE,

$$d\mathbf{Y}_t = b(\mathbf{Y}_t)dt + \sigma(\mathbf{Y}_t)d\mathbf{W}_t. \tag{3.4.1}$$

Here,

$$b(\mathbf{Y}_t) = \left(\frac{\theta_t^2}{2} \frac{\psi'(X_t)}{\psi(X_t)}, \quad \theta_t \left(q - \frac{\theta_t}{\sqrt{2\pi}} \frac{|\psi'(X_t)|}{\psi(X_t)} \right) \right)^T,$$

and

$$\sigma(\mathbf{Y_t}) = \begin{pmatrix} \theta_t & 0 \\ 0 & 0 \end{pmatrix},$$

2. Similarly the SMCMC converges weakly to a diffusion to the process which is the solution to the SDE

$$dX_t = \frac{\psi'(X_t)}{\psi(X_t)} \frac{\theta_0^2}{2} dt + \theta_0 dW_t. \tag{3.4.2}$$

and $\mathbf{W}_t$ is a two dimensional Brownian motion. See Remarks 12 for more details on the conditions on $\psi(\cdot)$. Here x^T is the transpose of a vector (or, a matrix) x.

Proof. Firstly, note that since $\mathbf{Y}_n(\frac{i}{n}) := \left(X_n(\frac{i}{n}), \quad \theta_n(\frac{i}{n})\right)$ is a homogeneous Markov chain it defines a transition kernel

$$\Pi_n(\mathbf{y}, A) = P\left(Y_n(\frac{i+1}{n}) \in A | \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y} \right), \quad \forall \mathbf{y} \in \mathbb{R} \times \mathbb{R}^+ \text{ and } \forall A \in \mathcal{B}(\mathbb{R} \times \mathbb{R}^+).$$

Note that since the initial points of the AMCMC and the SMCMC is fixed at (x_0, θ_0) Assumption 3 of Section 3.3 is automatically satisfied, where ν_0 is the degenerate distribution at (x_0, θ_0). The proof then follows essentially by obtaining the 'drift' and 'diffusion' coefficients of the discretized process, as in Equations (3.3.4) and then finding its limit. Formally, first obtain the quantities :

$$\begin{aligned} \mathbf{a}_n(\mathbf{y}, t) &:= (a_{n,i,j}(\mathbf{y}, t))_{i,j=1,2} := n \int_{\mathbb{R}} (\mathbf{z} - \mathbf{y})(\mathbf{z} - \mathbf{y})' \Pi_n(\mathbf{y}, d\mathbf{z}), \\ \mathbf{b}_n(\mathbf{y}, t) &:= (b_{n,k}(\mathbf{y}, t))_{k=1,2} := n \int_{\mathbb{R}} (\mathbf{z} - \mathbf{y}) \Pi_n(\mathbf{y}, d\mathbf{z}). \end{aligned}$$

The above is obtained by replacing h^{-1} by n in Equation (3.3.4).
Then find the matrix $\mathbf{a}$ and the vector $\mathbf{b}$ such that $\lim_{n\to\infty} ||\mathbf{a}_n(\mathbf{y},t) - \mathbf{a}(\mathbf{y},t)|| = 0$ and $\lim_{n\to\infty} ||\mathbf{b}_n(\mathbf{y},t) - \mathbf{b}(\mathbf{y},t)|| = 0$. Obtain the square root of matrix $\mathbf{a}(\mathbf{y},t)$(say $\sigma_(\mathbf{y},t)$), which satisfies $\mathbf{a}(\mathbf{y},t) = \sigma(\mathbf{y},t)\sigma(\mathbf{y},t)^T$. These coefficients define a diffusion process uniquely which is non-explosive (see Remark 12), and the limiting process is governed by the equation:

$$d\mathbf{Y}_t = \mathbf{b}(\mathbf{Y}_t,t)dt + \sigma(\mathbf{Y}_t,t)d\mathbf{W}_t,$$

where $\mathbf{W}_t$ is a two dimensional Wiener process. For the processes defined in (3.3.6) and (3.3.7), the limiting quantities $\mathbf{a}_n(\mathbf{y},t)$ and $\mathbf{b}_n(\mathbf{y},t)$ are
(for $\mathbf{y} = (x, \quad \theta)$):

$$\begin{aligned}
\lim_{n\to\infty} b_{n,1}(\mathbf{y},t) &= \frac{\theta^2}{2}\frac{\psi'(x)}{\psi(x)}, \\
\lim_{n\to\infty} b_{n,2}(\mathbf{y},t) &= \theta(q - \frac{\theta}{\sqrt{2\pi}}\frac{|\psi'(x)|}{\psi(x)}), \\
\lim_{n\to\infty} a_{n,1,1}(\mathbf{y},t) &= \theta^2, \\
\lim_{n\to\infty} a_{n,2,2}(\mathbf{y},t) &= 0, \\
\lim_{n\to\infty} a_{n,2,1}(\mathbf{y},t) &= 0 = \lim_{n\to\infty} a_{n,1,2}(\mathbf{y},t)
\end{aligned}$$

See Section 3.5 for the derivations.
Since the trace norm of a matrix is a continuous function of its components we can say that

$$||\mathbf{a}_n(\mathbf{y},t) - \mathbf{a}(\mathbf{y},t)|| \to 0 \text{ and } ||\mathbf{b}_n(\mathbf{y},t) - \mathbf{b}(\mathbf{y},t)|| \to 0$$

where

$$\begin{aligned}
\mathbf{a}(\mathbf{y},t) &= \begin{pmatrix} \theta^2 & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow \sigma(\mathbf{y},t) = \begin{pmatrix} \theta & 0 \\ 0 & 0 \end{pmatrix} \\
\text{and} \quad \mathbf{b}(\mathbf{y},t) &= \left(\frac{\theta^2}{2}\frac{\psi'(x)}{\psi(x)}, \quad \theta(q - \frac{\theta}{\sqrt{2\pi}}\frac{|\psi'(x)|}{\psi(x)}) \right)^T.
\end{aligned}$$

This proves the Theorem. ■

3.5 Drift and diffusion coefficients

Writing $\mathbf{y} = (x, \ \ \theta)$ we have

3.5.1 $b_{n,1}$

$$\begin{aligned}
& b_{n,1}(\mathbf{y}, t) \\
= \ & nE(X_n(\frac{i+1}{n}) - X_n(\frac{i}{n})| \ \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}), \quad \forall i = 0, 1, \ldots, \forall n \geq 1 \\
= \ & E(\sqrt{n}\theta_n(\frac{i}{n})\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})| \ \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) \\
= \ & \sqrt{n}\theta\Big(E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})I_{A_n}| \ X_n(\frac{i}{n}) = x, \ \theta_n(\frac{i}{n}) = \theta) \\
+ \ & E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})I_{A_n^c}| \ X_n(\frac{i}{n}) = x, \ \theta_n(\frac{i}{n}) = \theta)\Big).
\end{aligned}$$

where $A_n(= A_n(x, \theta))$ is the set where $\xi_n(\frac{i+1}{n})$ is one with probability 1, i.e,

$$\begin{aligned}
A_n(x, \theta) &= \{y : \frac{\psi(x + \frac{1}{\sqrt{n}}\theta y)}{\psi(x)} > 1\}. \\
\text{Thus,} \quad \lim_{n\to\infty} A_n^c(x, \theta) &= \begin{cases} (-\infty, 0) & \text{if } \psi'(x) > 0 \\ (0, \infty) & \text{if } \psi'(x) < 0. \end{cases}
\end{aligned}$$

Therefore,

$$
\begin{aligned}
b_{n,1}(\mathbf{y},t) &= \sqrt{n}\theta\Big(\int_{A_n}\epsilon\phi(\epsilon)d\epsilon+\int_{A_n^c}\frac{\psi(x+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)}\epsilon\phi(\epsilon)d\epsilon\Big)\\
&= \sqrt{n}\theta\Big(\int_{A_n}\epsilon\phi(\epsilon)d\epsilon+\int_{A_n^c}\epsilon\phi(\epsilon)d\epsilon\\
&+ \frac{\theta}{\sqrt{n}}\frac{\psi'(x)}{\psi(x)}\int_{A_n^c}\epsilon^2\phi(\epsilon)d\epsilon+O(\frac{1}{n})\Big),\quad\text{by Taylor's expansion,}\\
&= \sqrt{n}\theta\Big(\int_{\mathbb{R}}\epsilon\phi(\epsilon)d\epsilon+\frac{\theta}{\sqrt{n}}\frac{\psi'(x)}{\psi(x)}\int_{A_n^c}\epsilon^2\phi(\epsilon)d\epsilon+O(\frac{1}{n})\Big)\\
&= \theta^2\frac{\psi'(x)}{\psi(x)}\int_{A_n^c}\epsilon^2\phi(\epsilon)d\epsilon+O(\frac{1}{\sqrt{n}})\\
\Rightarrow\lim_{n\to\infty}b_{n,1}(\mathbf{y},t) &= \theta^2\frac{\psi'(x)}{\psi(x)}\lim_{n\to\infty}\int_{A_n^c}\epsilon^2\phi(\epsilon)d\epsilon\\
&= \begin{cases}\theta^2\frac{\psi'(x)}{\psi(x)}\int_{-\infty}^0\epsilon^2\phi(\epsilon)d\epsilon & \text{if } \psi'(x)>0\\ \theta^2\frac{\psi'(x)}{\psi(x)}\int_0^\infty\epsilon^2\phi(\epsilon)d\epsilon & \text{if } \psi'(x)<0\end{cases}\\
&= \frac{\theta^2}{2}\frac{\psi'(x)}{\psi(x)}.
\end{aligned}
$$

3.5.2 $b_{n,2}$

$$
\begin{aligned}
b_{n,2}(\mathbf{y},t) &= nE(\theta_n(\frac{i+1}{n})-\theta_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}),\quad\forall i=0,1,\dots\\
&= nE\Big(\theta_n(\frac{i}{n})\{e^{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))}-1\}|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&= n\theta\Big(\frac{1}{\sqrt{n}}E(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y})\\
&+ E(\frac{1}{2n}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y})+O(\frac{1}{n^{3/2}})\Big)\\
&= \theta\sqrt{n}E(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y})\\
&+ \frac{\theta}{2}E((\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y})+O(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

Now,

$$\begin{aligned}
& \theta\sqrt{n}E(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) \\
= \;& \theta\sqrt{n}\Big(E(\xi_n(\frac{i+1}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) - q_n(\frac{i}{n})\Big) \\
= \;& \theta\sqrt{n}\Big(\int_{A_n}\phi(\epsilon)d\epsilon + \int_{A_n^c}\frac{\psi(x+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)}\phi(\epsilon)d\epsilon - q_n(\frac{i}{n})\Big) \\
= \;& \theta\sqrt{n}\Big(\int_{A_n}\phi(\epsilon)d\epsilon \\
+ \;& \int_{A_n^c}\{1+\frac{\theta}{\sqrt{n}}\frac{\psi'(x)}{\psi(x)}\epsilon + O(\frac{1}{n})\}\phi(\epsilon)d\epsilon - q_n(\frac{i}{n})\Big) \\
= \;& \theta\sqrt{n}(1-q_n(\frac{i}{n})) \\
+ \;& \theta^2\frac{\psi'(x)}{\psi(x)}\int_{A_n^c}\epsilon\phi(\epsilon)d\epsilon + O(\frac{1}{\sqrt{n}}).
\end{aligned} \tag{3.5.3}$$

And,

$$\begin{aligned}
& E\Big((\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
= \;& E\Big(\xi_n(\frac{i+1}{n})^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
- \;& 2q_n(\frac{i}{n})E\Big(\xi_n(\frac{i+1}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) + q_n(\frac{i}{n})^2 \\
= \;& \int_{A_n}\phi(\epsilon)d\epsilon + \int_{A_n^c}\frac{\psi(x+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)}\phi(\epsilon)d\epsilon \\
- \;& 2q_n(\frac{i}{n})\Big(\int_{A_n}\phi(\epsilon)d\epsilon + \int_{A_n^c}\frac{\psi(x+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)}\phi(\epsilon)d\epsilon\Big) \\
+ \;& q_n(\frac{i}{n})^2 \\
= \;& (1-q_n(\frac{i}{n}))^2 + \frac{1}{\sqrt{n}}(1-2q_n(\frac{i}{n}))\theta\frac{\psi'(x)}{\psi(x)}\int_{A_n^c}\epsilon\phi(\epsilon)d\epsilon \\
+ \;& O(\frac{1}{n}) \longrightarrow 0,
\end{aligned} \tag{3.5.4}$$

as $n \to \infty$ (since $1 - q_n(\frac{i}{n}) \approx \frac{q}{\sqrt{n}}$), therefore

$$\frac{1}{\sqrt{n}}(1-2q_n(\frac{i}{n})) \;\approx\; \frac{1}{\sqrt{n}}(\frac{2q}{\sqrt{n}} - 1).$$

Thus, from (3.5.3) and (3.5.4) we have,

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{b}_{n,2}(\mathbf{y},t) &= \theta q + \theta^2 \frac{\psi'(x)}{\psi(x)} \lim_{n\to\infty} \int_{A_n^c} \epsilon\phi(\epsilon)d\epsilon \\
&= \begin{cases} \theta\Big(q + \frac{\theta}{\sqrt{2\pi}}\frac{\psi'(x)}{\psi(x)}\Big) & \text{if } \psi'(x) < 0 \\ \theta\Big(q - \frac{\theta}{\sqrt{2\pi}}\frac{\psi'(x)}{\psi(x)}\Big) & \text{if } \psi'(x) > 0 \end{cases} \\
&= \theta\Big(q - \frac{\theta}{\sqrt{2\pi}}\frac{|\psi'(x)|}{\psi(x)}\Big).
\end{aligned}$$

3.5.3 $a_{n,1,1}$.

$$\begin{aligned}
a_{n,1,1}(\mathbf{y},t) &= nE\Big((X_n(\frac{i+1}{n}) - X_n(\frac{i}{n})^2)|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \ \forall i = 0,1,\ldots \\
&= \theta^2 E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) \\
&= \theta^2\Big(E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})^2 I_{A_n}|\ \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) \\
&+ E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})^2 I_{A_n^c}|\ \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y})\Big) \\
&= \theta^2\Big(\int_{A_n} \epsilon^2\phi(\epsilon)d\epsilon + \int_{A_n^c} \epsilon^2 \frac{\psi(x + \frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)}\phi(\epsilon)d\epsilon\Big) \\
&= \theta^2\Big(\int_{A_n} \epsilon^2\phi(\epsilon)d\epsilon + \int_{A_n^c} \epsilon^2\phi(\epsilon)d\epsilon + O(\frac{1}{\sqrt{n}})\Big) \\
&= \theta^2 + O(\frac{1}{\sqrt{n}}). \\
\Rightarrow \lim_{n\to\infty} a_{n,1,1}(t) &= \theta^2.
\end{aligned}$$

3.5.4 $a_{n,2,2}$.

$$
\begin{aligned}
a_{n,2,2}(\mathbf{y},t) &= nE\Big((\theta_n(\frac{i+1}{n})-\theta_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&= nE\Big(\theta_n(\frac{i}{n})^2(e^{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))}-1)^2|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&= n\theta^2E\Big(\Big\{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))+\frac{1}{2n}(\xi_n(\frac{i+1}{n})\\
&- q_n(\frac{i}{n}))^2+O(\frac{1}{n^{3/2}})\Big\}^2|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&= \theta^2E\Big((\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&+ O(\frac{1}{\sqrt{n}})\\
\Rightarrow \lim_{n\to\infty} a_{n,2,2}(\mathbf{y},t) &= \theta^2\lim_{n\to\infty}E\Big(\Big(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n})\Big)^2|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)=0,
\end{aligned}
$$

from (3.5.4).

3.5.5 $a_{n,1,2}$ and $a_{n,2,1}$.

$$
\begin{aligned}
a_{n,1,2}(\mathbf{y},t) &= nE\Big(\{X_n(\frac{i+1}{n})-X_n(\frac{i}{n})\}\{\theta_n(\frac{i+1}{n})-\theta_n(\frac{i}{n})\}|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&= nE\Big(\{\frac{1}{\sqrt{n}}\theta_n(\frac{i}{n})\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})\}\{\theta_n(\frac{i}{n})(e^{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))}-1)\}\Big)\\
&= \sqrt{n}\theta^2E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})\Big\{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))\\
&+ O(\frac{1}{n})\Big\}|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&= \theta^2E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&+ O(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

Since $\xi_n = 0$, or 1, $\xi_n^2 = \xi_n$. Hence $\xi_n\epsilon_n(\xi_n - q_n) = \xi_n^2\epsilon_n - \xi_n\epsilon_n q_n = \xi_n\epsilon_n(1-q_n)$. Therefore,

$$\begin{aligned} & E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\ = \;& (1 - q_n(\frac{i}{n}))E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\ = \;& (1 - q_n(\frac{i}{n}))O(1) \;\; \longrightarrow 0, \quad \text{as} \quad n \to \infty. \end{aligned}$$

Thus, $\lim_{n\to\infty} a_{n,1,2} = \lim_{n\to\infty} a_{n,2,1} = 0$. ■

Remark 11. Note that the form of the SDE in Theorem 3 is similar to the Langevin diffusion equation (for univariate densities). This shows that this adaptive MCMC properly Normalized behaves in the limit as the Langevin diffusion which has $\psi(\cdot)$ as the invariant distribution. This bears a little resemblance to the Metropolis adjusted Langevin algorithm (MALA) procedure, where the proposal emulates the discretization of the Langevin algorithm. For more information regarding MALA and its convergence properties, see Marshall and Roberts [42], Roberts and Rosenthal [58].

Remark 12. For a general target distribution $\psi(\cdot)$ we assume that the solutions satisfy the non-explosive condition given by, see Skorohod ([62])

$$|b(\mathbf{y},t)| + |\sigma(\mathbf{y},t)| \;\; \leq \;\; C(1 + |\mathbf{y}|), \tag{3.5.5}$$

for some constant $C > 0$. We also assume that it also satisfies the local Lipschitz condition for uniqueness given by

$$|b(\mathbf{y}_1,t) - b(\mathbf{y}_2,t)| + ||\sigma(\mathbf{y}_1,t) - \sigma(\mathbf{y}_2,t)|| \;\; \leq \;\; D_k(|\mathbf{y}_1 - \mathbf{y}_2|, \tag{3.5.6}$$

where $\mathbf{y}_1, \mathbf{y}_2$ lies in some compact interval $S_k \subset \mathbb{R} \times \mathbb{R}^+$ and some constant $D_k > 0$. Here

$$||\sigma(\mathbf{y},t)|| = \sqrt{\sum_{i,j=1}^{2} \sigma_{i,j}^2}.$$

For constant θ_t the non-explosive condition boils down to

$$\frac{|\psi'(x)|}{\psi(x)} \;\; \leq \;\; C(1 + |x|), \tag{3.5.7}$$

for some $C \geq 0$.

REMARK 13. If the target density $\psi(\cdot)$ satisfy the linear growth condition

$$\frac{|\psi'(x)|}{\psi(x)} \leq a|x| + b,$$

for some $a, b > 0$, then from the SDE (3.4.1) we have that

$$\begin{aligned} d\theta_s &= \theta_s\Big(q - \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_s)|}{\psi(X_s)}\Big)ds \leq q\theta_s ds \\ \Rightarrow \theta_s &\leq \theta_0 e^{qs} \quad \text{and,} \\ dX_s &= \frac{\theta_s^2}{2}\frac{\psi'(X_s)}{\psi(X_s)} + \theta_s dW_s. \end{aligned}$$

Taking integrals from 0 to t we have

$$\begin{aligned} X_t &= X_0 + \int_0^t \frac{\theta_s^2}{2}\frac{\psi'(X_s)}{\psi(X_s)}ds + \int_0^t \theta_s dW_s \\ \Rightarrow |X_t| &\leq |X_0| + \int_0^t \frac{\theta_s^2}{2}\frac{|\psi'(X_s)|}{\psi(X_s)}ds + |\int_0^t \theta_s dW_s| \\ &\leq |X_0| + \int_0^t \frac{\theta_s^2}{2}\Big(a|X_s| + b\Big)ds + |\int_0^t \theta_s dW_s| \\ &\leq |X_0| + \frac{a\theta_0^2}{2}\int_0^t e^{2qs}|X_s|ds + \frac{b\theta_0^2}{4q}(e^{2qt} - 1) + |\int_0^t \theta_s dW_s|, \end{aligned}$$

using the bound for θ_t. Taking expectations we have

$$\Rightarrow E(|X_t|) \leq E(|X_0|) + \frac{aE(\theta_0^2)}{2}\int_0^t e^{2qs}E(|X_s|)ds + \frac{bE(\theta_0^2)}{4q}(e^{2qt} - 1) + E(|\int_0^t \theta_s dW_s|).$$

By the Cauchy Schwarz inequality the last expectation is bounded by

$$\sqrt{E(\int_0^t \theta_s dW_s)^2} = \sqrt{\int_0^t E(\theta_s^2)ds} \leq \sqrt{E(\theta_0^2)}\sqrt{\int_0^t e^{2qs}ds} = \frac{\sqrt{E(\theta_0^2)}}{\sqrt{2q}}\sqrt{e^{2qt} - 1} \leq \frac{\sqrt{E(\theta_0^2)}}{\sqrt{2q}}e^{qt}.$$

Hence by a rearrangement of terms we have

$$E|X_t| \leq \underbrace{E(|X_0|) + \frac{bE(\theta_0^2)}{4q}(e^{2qt}-1) + \frac{\sqrt{E(\theta_0^2)}}{\sqrt{2q}}e^{qt}}_{F_t} + \int_0^t \underbrace{\frac{aE(\theta_0^2)}{2}e^{2qs}}_{A_s} E(|X_s|)ds$$

Writing $G_t = E|X_t|,\ F_t = E(|X_0|) + \frac{bE(\theta_0^2)}{4q}e^{2qt} + \frac{\sqrt{E(\theta_0^2)}}{\sqrt{2q}}e^{qt},\ A_t = \frac{aE(\theta_0^2)e^{2qt}}{2}$ we have

$$G_t \leq F_t + \int_0^t A_s G_s ds,$$

where F_t is non negative and A_t is increasing as a function of $t \in [0,\infty)$. Therefore from Gronwall's inequality, see, for example, [51], pp. 78, we have

$$G_t \leq F_t e^{\int_0^t A_s ds},\ t \geq 0.$$

Now

$$\int_0^t A_s ds = \frac{aE(\theta_0^2)}{4q}(e^{2qt}-1) \leq \frac{aE(\theta_0^2)}{4q}e^{2qt},$$

and so

$$E|X_t| \leq \left(E(|X_0|) + \frac{bE(\theta_0^2)}{4q}e^{2qt} + \frac{\sqrt{E(\theta_0^2)}}{\sqrt{2q}}e^{qt}\right)e^{\frac{aE(\theta_0^2)}{4q}e^{2qt}}.$$

This proves that the solution to the SDE of (X_t, θ_t) given by Equation (3.4.1) is non-explosive.

3.6 Comparison between Adaptive and non-adaptive MCMC by simulations

In this section we compare both the discrete and continuous non-adaptive (also called standard) MH sampler against its adaptive counterpart as proposed in Section 3.2. We try to simulate samples from target distributions with heavier and lighter tails compared to the Normal distribution. In the standard MH the tuning parameter is kept fixed. Since the tuning parameter that gives the best result is not known in general, we compare the simulations for a multitude of θ values. (Note that results relating to optimal q for Normal distribution is given in Gelman *et al.* [25] which have been extended further in Bedard and Rosenthal [9]).

3.6.1 Comparison between the discrete time chains

We generate the discrete time version of the adaptive and non-adaptive sampler for different values of q and the starting value, θ_0. All the proposal distributions are Normal and the target density is Normal(0,1) in Table 3.1 and Cauchy(0,1) in Table 3.2 . After generating a sample of size 10000 we discard the first 1000 samples as burn-in. To check the efficiency of the sampler we perform the one sample Kolmogorov-Smirnov (KS) test on the remaining sample and find the asymptotic q value of the KS test statistic D measuring the distance between the empirical distribution of the generated sample and the target distribution.

A measure of the amount of mixing is **Expected Square Jumping Distance**(ESJD) defined as $E(X_i - X_{i-1})^2$. Based on the generated sample it can be estimated by $E := \frac{1}{n-B} \sum_{i=B+1}^{n} (X_i - X_{i-1})^2$, where B is the size of burn-in sample. In general, higher value of ESJD implies greater mixing , (also see Gelman and Pasarica [26] for more details).

See Tables 3.1, 3.2 here.

From Table 3.1 we see that the starting value $\theta_0 = 2.38$ is the best choice with respect to the non-Adaptive chain as well as the Adaptive chain. This value of θ_0 was suggested by Gelman *et al.* [25]. For adaptive chain the optimal value of q lies somewhere between 0.25 and 0.50. Again, by [25], the optimal value of acceptance probability was close to 0.238. Therefore our simulations corroborates their findings to some extent.

It is well known that the naive MH algorithm is not efficient enough in simulating from a Heavy tailed distribution. This is what we see in Table 3.2). But we see that for the

adaptive version with q lying in the same interval, i.e., in [0.25,0.50], its performance is at least better than that of its non-adaptive counterpart, although by itself it is not quite efficient.
In comparison we see that the continuous time version of the AMCMC performs much better than its discrete counterpart as elucidated in the next section.

3.6.2 Comparison between the continuous time processes

To compare how fast the two diffusion processes which are solution to (3.4.1) and (3.4.2) converge to stationarity we apply Euler discretization to each of them for various choices of target density ψ and mesh size h. For the process (3.4.1) the Euler discretization is given by $X_{ih}, \quad i = 0, 1, 2 \ldots, \frac{T}{h}$ where:

$$\begin{aligned} X_{(i+1)h} &= X_{ih} + h\frac{\psi'(X_{ih})}{\psi(X_{ih})}\frac{\theta_{ih}^2}{2} + \sqrt{h}\theta_{ih}Z_{ih}^{(1)} \\ \theta_{(i+1)h} &= \theta_{ih} + h\Big(\theta_{ih}(q - \theta_{ih}\frac{|X_{ih}|}{\sqrt{2\pi}})\Big), \quad i = 0, 1, \ldots, T/h. \end{aligned}$$

Similarly the Euler discretization for the process (3.4.2) is $\{Y_{ih}\}$ where:

$$Y_{(i+1)h} = Y_{ih} + h\frac{\psi'(X_{ih})}{\psi(X_{ih})}\frac{\theta^2}{2} + \sqrt{h}\theta Z_{ih}^{(2)}, \quad i = 0, 1, \ldots, T/h, \ \theta = \theta_0 \in \mathbb{R}^+$$

Where $Z_{ih}^{(j)}, \quad i = 0, 1, \ldots, T/h, \quad j = 1, 2$ are independent $N(0, 1)$ random variables.

For various values of the mesh size h we simulate 1000 parallel SDE using the Euler discretization for the Adaptive and the Standard MCMC and obtain the value of X_T at time $T = 1$. Table 3.3, Table 3.4 and Table 3.5 give the result when the target distribution are $N(0, 1)$, Cauchy$(0, 1)$ and Exponential(1) distribution, respectively. We also compute the Kolmogorov Smirnov distance between the sample and the target distribution and also find its asymptotic p value.

See Table 3.3, 3.4 here, 3.5.

It should be noted that since the support of the Exponential density (Table 3.5) is only the positive part of the real line, while simulating the SDE corresponding to the distribution any move to the left of zero was modified accordingly.

The tables clearly indicate that for a proper choice of the parameter q, the Adaptive version performs better than the non-Adaptive version. Almost always the asymptotic p-value is small for smaller values of q it reaches its peak at an optimum q and then decreases.The reason is that if q is small then the quantity $-\theta_{ih}|X_{ih}|/\sqrt{2\pi}$ dominates q and θ_{ih} decreases on the average. On the other hand if q is large then q dominates and θ_{ih} increases on the average. As a result the sample thus generated differs widely from the target density.

Another table of interest is Table 3.4. Standard MCMC is not quite adept in sampling from a density with heavy tails. AMCMC to some degree addresses this problem where we see that the p-value is always higher for the Adaptive case for all values of mesh size h.

In Figures 3.2 and 3.4 we give simulation of sample paths of discrete AMCMC and SMCMC together with the plot of θ_t when the target is standard Normal and Cauchy respectively.

3.7 Summary

Diffusion approximation is a well studied technique that has been applied to many fields (e.g., [22], [48]). In AMCMC the tuning parameter changes as the iteration progresses and therefore the transition kernel also changes. As a result the invariant properties of the chain are not easily obtainable. In this chapter, we have applied the diffusion approximation procedure to the AMCMC chain and obtained the limiting SDE to arrive at the target distribution. Diffusive limits for Metropolis Hastings algorithm were earlier

		Adpative			non-Adaptive		
θ_0	q	D	p-value	E	D	p-value	E
	0.10	0.0502	2.2e-12	0.2388			
	0.25	0.0278	1.729e-6	0.6064			
0.10	0.50	0.0165	0.01472	0.7150	0.1369	2.2e-16	0.00934
	0.75	0.025	2.702e-5	0.3709			
	0.10	0.0391	2.244e-12	0.2587			
	0.25	0.0258	1.199e-5	0.5926			
0.25	0.50	0.024	6.507e-5	0.6971	0.0377	1.56e-11	0.05228
	0.75	0.0273	3.006e-6	0.3618			
	0.10	0.0458	2.2e-16	0.2356			
	0.25	0.0186	0.003867	0.5835			
1.0	0.50	0.0168	0.01270	0.7137	0.0251	2.29e-5	0.46209
	0.75	0.0272	3.344e-6	0.3642			
	0.10	0.039	2.46e-16	0.2572			
	0.25	0.0138	0.06627	0.6050			
2.38	0.50	0.0153	0.02883	0.7070	0.0223	2.69e-4	0.71047
	0.75	0.0233	0.00010	0.3648			
	0.10	0.0467	1.066e-14	0.0427			
	0.25	0.0374	2.38e-11	0.5953			
10	0.50	0.0221	3.012e-4	0.6988	0.0302	1.444e-7	0.26411
	0.75	0.0208	0.00081	0.3613			
	0.10	0.0467	2.2e-16	0.2433			
	0.25	0.0272	3.42e-16	0.5894			
20	0.50	0.030	1.361e-7	0.6992	0.0669	2.26e-15	0.14603
	0.75	0.0225	0.002141	0.3709			

TABLE 3.1: Table comparing the asymptotic p-values of sample generated using AMCMC and SMCMC for different values of q and θ_0 where the target density is Normal(0,1)

obtained in [58, 63, 64]. Also, there are some recent work on diffusive limits of high-dimensional non-adaptive MCMC that came to our attention (for example, see Mattingly *et al.* [43]). However, to the best of our knowledge, application of diffusion approximation to Adaptive MCMC and subsequent comparison between AMCMC and Standard MCMC using their respective diffusive limits have not been done earlier. Our technique expands the scope of comparison between AMCMC and Standard MCMC, as embedding in continuous time allows various discrete approximations through which one can compare them in finer details.

		Adaptive			non-Adaptive		
θ_0	q	D	p-value	E	D	p-value	E
	0.10	0.0675	2.2e-16	34.8595			
	0.234	0.0246	3.65e-5	10.6793			
0.10	0.50	0.0272	3.419e-6	2.9033	0.2023	2.2e-16	0.00942
	0.75	0.0335	3.299e-9	0.7969			
	0.10	0.0582	2.2e-16	81.86962			
	0.234	0.0389	2.79e-12	10.3054			
0.25	0.50	0.0186	0.00399	2.9030	0.1069	2.2e-16	0.5468
	0.75	0.0385	5.29e-12	0.7305			
	0.10	0.0455	2.2e-16	34.8598			
	0.234	0.0332	4.601e-9	10.7424			
1.0	0.50	0.0227	1.938e-4	2.9033	0.0418	4.47e-14	0.6453
	0.75	0.0336	2.81e-9	0.7144			
	0.10	0.0675	2.2e-16	34.8598			
	0.234	0.024	6.175e-5	11.0824			
2.38	0.50	0.0213	5.7137e-4	2.7663	0.0302	1.493e-7	1.01557
	0.75	0.0353	3.646e-10	0.7490			
	0.10	0.0528	2.2e-16	34.8598			
	0.234	0.0383	6.56e-12	10.0612			
10	0.50	0.0192	0.00258	2.8593	0.0334	4.028e-9	10.19617
	0.75	0.0318	2.538e-8	0.7582			
	0.10	0.0529	2.2e-16	19.5438			
	0.234	0.034	1.836e-9	10.4076			
20	0.50	0.0224	3.831e-5	2.8915	0.0415	6.40e-14	22.9938
	0.75	0.0306	9.313e-8	0.7419			

TABLE 3.2: Table comparing the asymptotic p values of sample generated using AMCMC and SMCMC for different values of q and θ_0 where the target density is Cauchy(0,1)

q	p value		$\theta(T)$	D	
	Adaptive	non-Adaptive		Adaptive	non-Adaptive
h=0.0001					
1.0	0.0385		4.5900	0.0444	
2.0	**0.2035**	**0.5273**	**0.0338**	**0.0338**	**0.0256**
2.5	0.1415		8.8963	0.0364	
h=0.0005					
4.5	0.4681		12.8575	0.0268	
5.0	**0.4774**	**0.28**	**14.4240**	**0.0266**	**0.0313**
5.5	0.4186		15.9708	0.0279	
6.0	0.2254		17.50510	0.033	0.0313
h=0.001					
4	0.1999		15.7125	0.0305	
5	0.3804		14.0184	0.02870	
5.5	0.3409		20.9751	0.0297	
6.0	0.4179		17.6700	0.0279	
6.5	**0.4393**	**0.628**	**18.9813**	**0.0274**	**0.0237**
7.5	0.3369		21.6681	0.0298	
8.0	0.2127		23.0411	0.0335	
h=0.005					
0.2	0.2900		2.4322	0.0310	
0.5	0.2595		2.7888	0.0319	
1.0	0.1761		3.5346	0.0348	
1.5	**0.4378**	**6.08e-10**	**4.5760**	**0.0275**	**0.096**
2.0	0.3782		5.9303	0.0288	
2.5	0.2489		7.3848	0.0323	
3.0	0.08535		8.8085	0.0397	
h=0.01					
1.0	0.0033		3.6390	0.0565	
1.5	0.0180		5.0818	0.0485	
2.0	**0.0977**	**0**	**6.4826**	**0.0388**	**0.4516**.
3.0	0.0347		9.2409	0.0450	

TABLE 3.3: Simulation of the SDE for Normal target density

q	p value		$\theta(T)$	D	
	Adaptive	non-Adaptive		Adaptive	non-Adaptive
h=0.0001					
5	0.2938		18.13893	0.0309	
7	**0.4407**	**0.7756**	**25.15494**	**0.0274**	**0.0209**
8	0.3306		27.8692	0.03	
h=0.0005					
3.0	1.10e-11		14.5622	0.1139	
4.0	0.07		22.2738	0.0409	
4.5	0.1104		26.3661	0.0381	
6.0	**0.2259**	**0.4709**	**35.3313**	**0.033**	**0.0268**
7.0	0.2022		41.4020	0.0338	
h=0.001					
0.5	0.1423		4.2115	0.0363	
5	0.168		18.5107	0.0352	
6	**0.2353**	**0.4894**	**21.3778**	**0.0327**	**0.0264**
7	0.1641		24.8888	0.0354	
h=0.005					
2	0.0356		10.0997	0.0457	
2.5	0.0481		9.9744	0.0481	
3	0.0824		18.6841	0.0399	
3.5	**0.1197**	**1.652e-6**	**24.7255**	**0.0375**	**0.0837**
4.0	0.0627		31.845	0.0416	
h=0.01					
0.5	0.0762		3.8347	0.0404	
1.0	0.0162		4.8314	0.0491	
2.0	0.0277		8.1943	0.0462	
2.5	0.0060		9.9910	0.0539	
2.75	**0.0129**	**0**	**10.7327**	**0.0502**	**0.1537**
3.0	0.0079		11.2480	0.0526	
3.5	0.0004		12.9623	0.06491	

TABLE 3.4: Simulation of the SDE for Cauchy target density

q	p value		$\theta(T)$	D	
	Adaptive	non-Adaptive		Adaptive	non-Adaptive
h=0.0001					
1.5	0.1885		4.5917	0.0344	
2.0	0.4301		5.5788	0.0276	
2.5	**0.4880**	**0.3641**	**6.6407**	**0.0264**	**0.0291**
3.0	0.3783		7.7608	0.0288	
20	0.0507		50.13257	0.0429	
h=0.0005					
2.0	0.02508		5.5782	0.0468	
2.5	0.07267		6.6402	0.0407	
3.0	0.1024		7.7604	0.0385	
3.5	0.2379		8.9240	0.0326	
4.0	**0.7415**	**0.6055**	**10.1186**	**0.0216**	**0.0241**
4.25	0.4527		10.7244	0.0271	
4.5	0.04414		11.3345	0.0274	
h=0.001					
1.5	0.6933		4.5905	0.0225	
2.0	**0.9175**	**0.3688**	**5.5774**	**0.0176**	**0.029**
2.5	0.5089		6.6394	0.026	
h=0.005					
5.5	0.328		13.80297	0.03	
6.0	**0.9579**	**0.893**	**15.0483**	**0.0161**	**0.0183**
6.5	0.4649		16.2972	0.0269	
h=0.01					
6.0	0.6368		15.0477	0.0235	
6.5	**0.9041**	**0.4622**	**16.2969**	**0.0179**	**0.0269**
7	0.05136	0.4622	17.548	0.0428	

TABLE 3.5: Simulation of the SDE for exponential target density

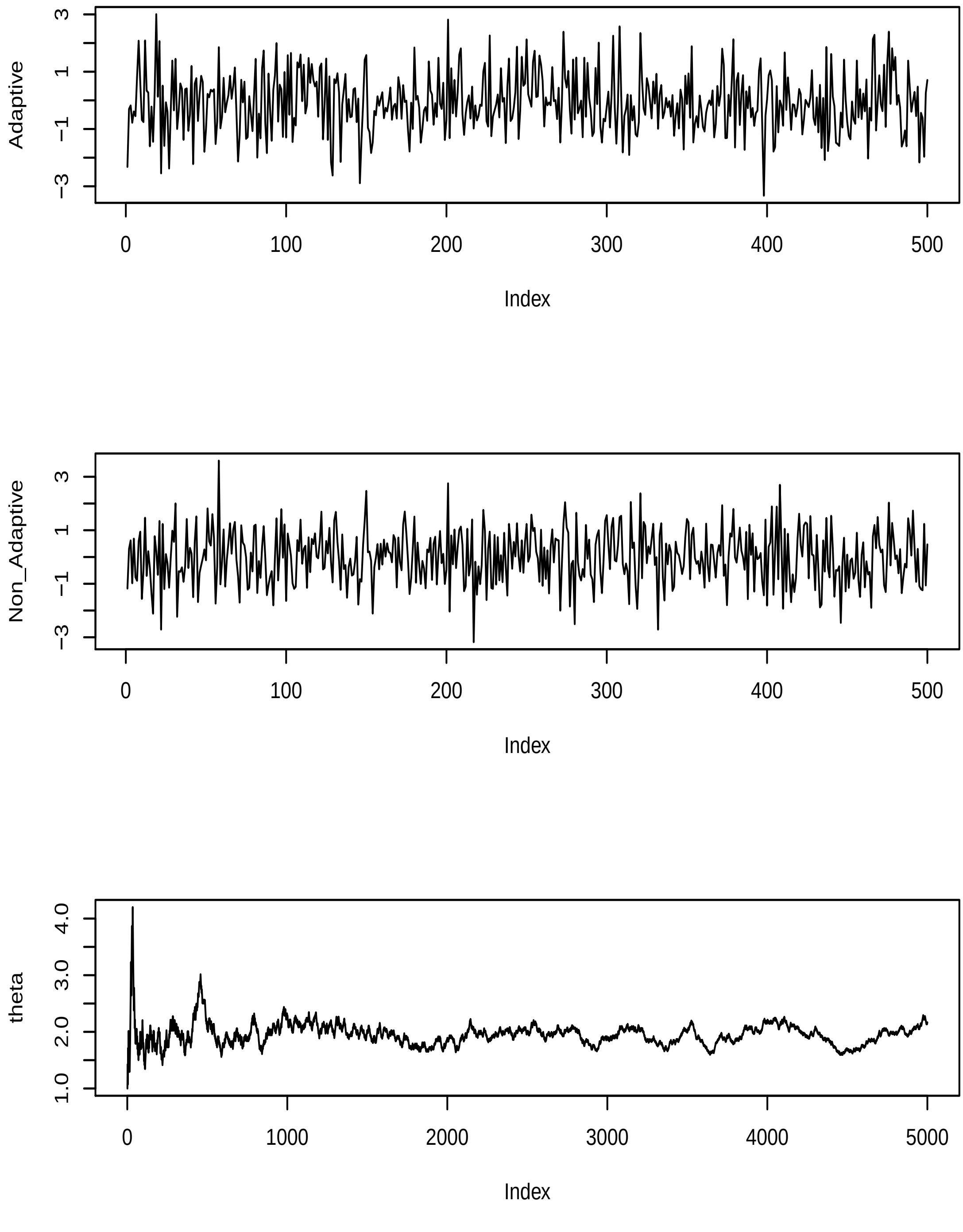

FIGURE 3.1: Discrete AMCMC and SMCMC plot for Normal(0,1) with q=0.50 and $\theta_0 = 1$.

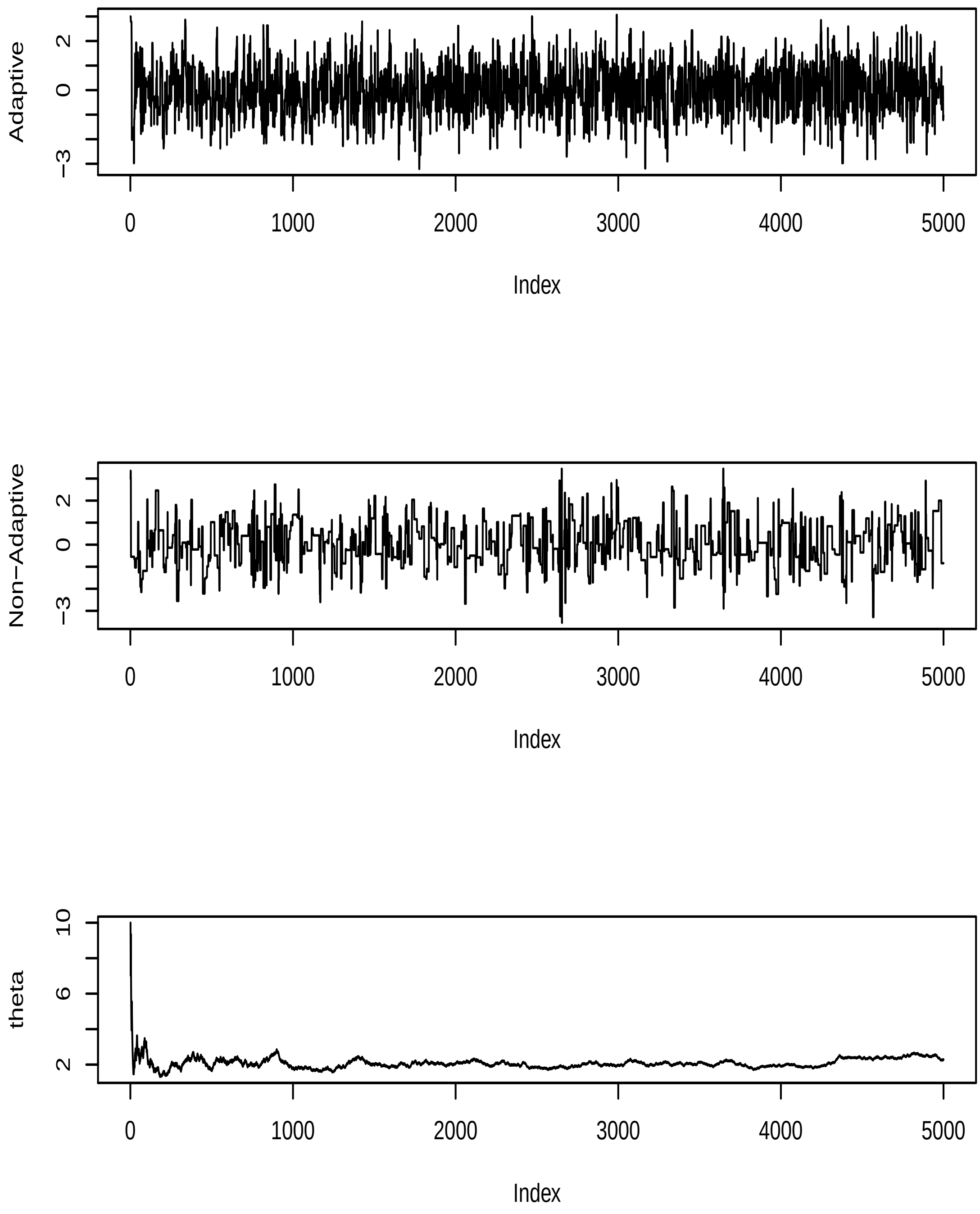

FIGURE 3.2: Discrete AMCMC and SMCMC plot for Normal(0,1) with q=0.75 and $\theta_0 = 10$.

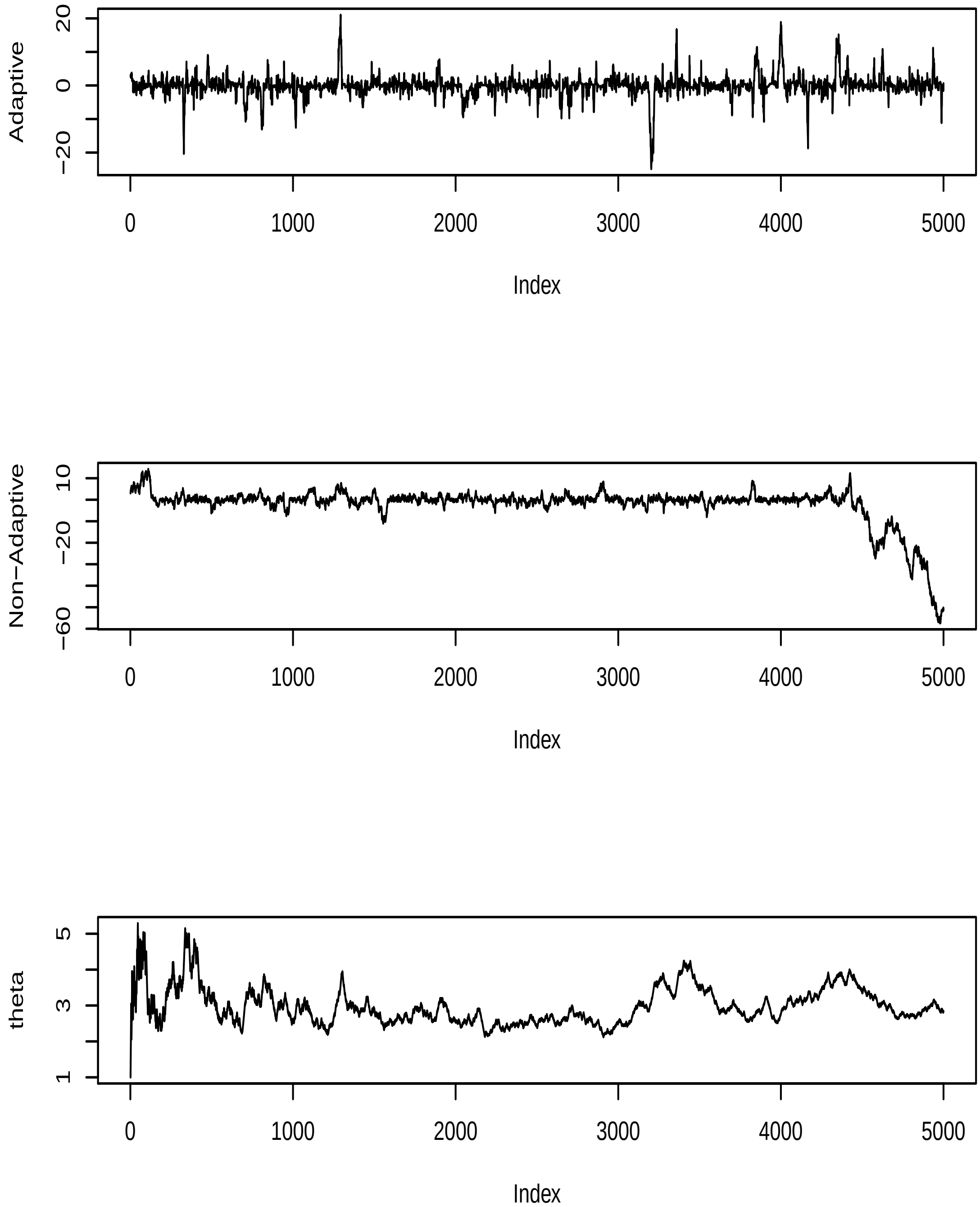

FIGURE 3.3: Discrete AMCMC and SMCMC plot for Cauchy(0,1) with q=0.50 and $\theta_0 = 1$.

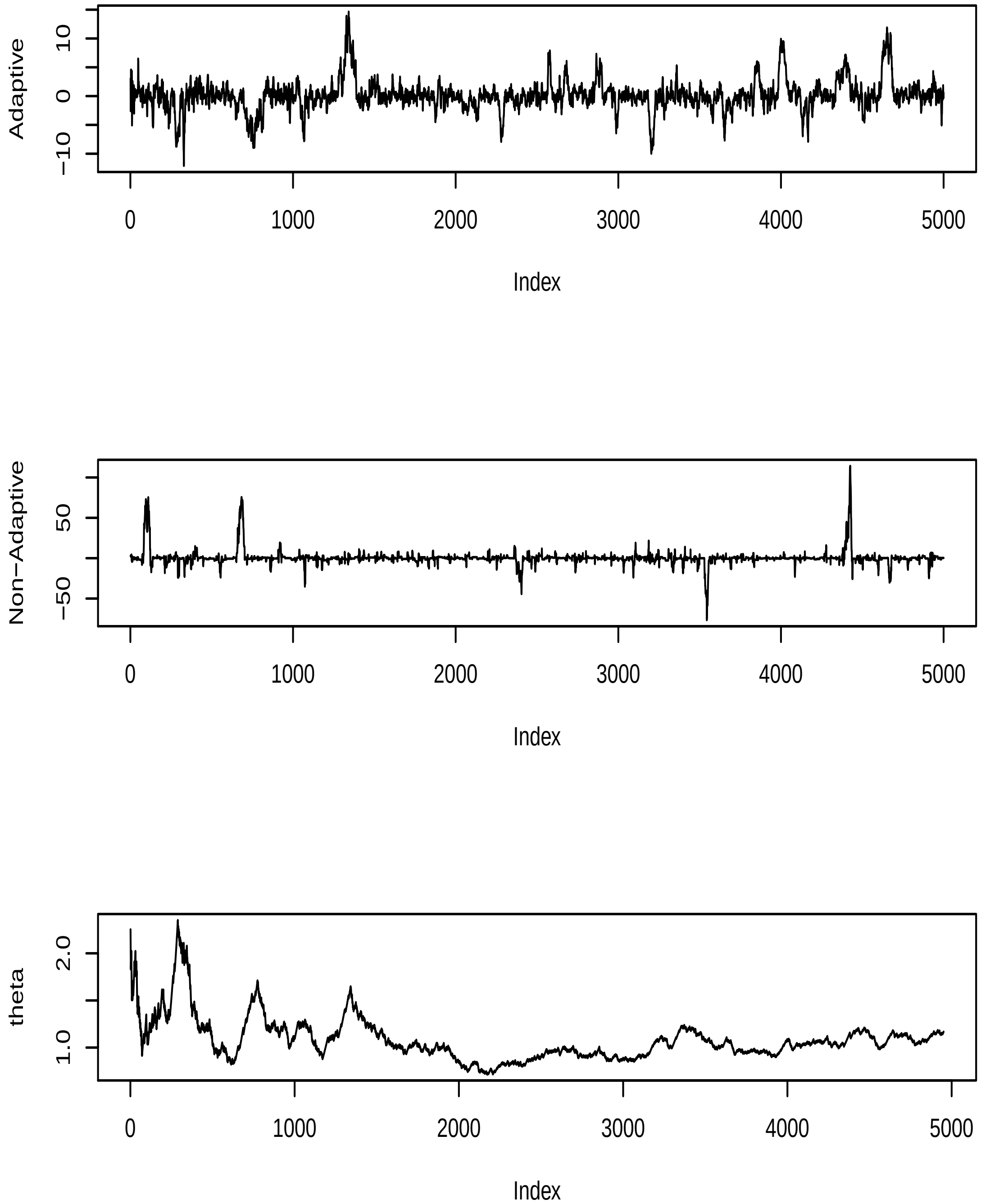

FIGURE 3.4: Discrete AMCMC and SMCMC plot for Cauchy(0,1) with q=0.75 and $\theta_0 = 10$.

Chapter 4

Diffusive limits when the target distribution is Standard Normal

4.1 Introduction

In the last chapter, using the Diffusion Approximation technique we observed that the limiting process of the AMCMC is governed by the following Theorem:

Proposition 14. (from Theorem 3 of Chapter 3) The limit of the process $\mathbf{Y}_n(t) := \Big(X_n(t),\ \theta_n(t)\Big)'$, where $X_n(t)$ and $\theta_n(t)$ is given by (3.3.6) and (3.3.7) respectively, is governed by the SDE:

$$d\mathbf{Y}_t \quad = \quad b(\mathbf{Y}_t)dt + \sigma(\mathbf{Y}_t)d\mathbf{W}_t, \quad \text{with} \;\; \mathbf{Y}_t = (X_t, \theta_t)',$$

where,

$$b(\mathbf{Y}_t) = \left(\frac{\theta_t^2}{2}\frac{\psi'(X_t)}{\psi(X_t)},\ \ \theta_t\left(q - \frac{\theta_t}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}\right)\right)',$$

$$\sigma(\mathbf{Y_t}) = \begin{pmatrix} \theta_t & 0 \\ 0 & 0 \end{pmatrix}$$

and $\mathbf{W}_t$ is a two dimensional Wiener process and $\frac{\psi'(\cdot)}{\psi(\cdot)}$ satisfies the linear growth condition given in Remark 13.

This chapter is arranged as follows. In Section 4.2.1 we show that the process is tight. This combined with the hypoelliptic condition in Section 4.2.2 shows that the process admits a smooth invariant distribution. After establishing moment conditions of the variables under consideration in Section 4.2.1.1 and Section 4.2.1.4, identification of the target distribution is proved in Section 4.2.3.

4.2 Main result

In this chapter we concentrate on the case where the target density is standard Normal (i.e., $\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$). Then the SDE takes the form:

$$\begin{aligned} d\mathbf{Y}_t &= b(\mathbf{Y}_t)dt + \sigma(\mathbf{Y}_t)d\mathbf{W}_t, \qquad \text{where,} \\ b(\mathbf{Y}_t) &= \left(-\frac{\theta_t^2}{2}X_t, \;\; \theta_t\left(q - \frac{\theta_t}{\sqrt{2\pi}}|X_t|\right)\right)'. \end{aligned} \tag{4.2.1}$$

and $\sigma(\mathbf{Y}_t)$ remains the same. Throughout the chapter we assume $\mathbf{Y}_0 = (X_0, \theta_0)'$ is independent of $\{W_t : t \geq 0\}$.

REMARK 14. Equation (4.2.1) when written in a more explicit form becomes :

$$\begin{aligned} dX_t &= -X_t\frac{\theta_t^2}{2} + \theta_t dW_t \\ d\theta_t &= \theta_t\Big(q - \frac{\theta_t}{\sqrt{2\pi}}|X_t|\Big)dt \end{aligned}$$

It resembles that of a coupled Ornstein Uhlenbeck (OU) process with the diffusion coefficient itself following a logistic equation. One knows that for a standard OU process the N(0,1) distribution is the invariant distribution. In the above case, it is slightly complicated since the diffusion coefficient is not constant. We show that even then the limiting distribution of the diffusion process is Normal.

REMARK 15. It will be shown in Lemma 14 that $E(X_t^2) < \infty, \;\forall t > 0$. This implies that $X_t < \infty$ a.s $\forall t$. From the SDE of θ_t it is shown (see Equation (4.2.2)) that $\theta_t \leq \theta_0 e^{qt}$.

Consequently $\theta_t < \infty$ almost surely. Therefore the solutions of Equation (4.2.1) is non-explosive.

Here is the main Theorem of this chapter:

THEOREM 4. The X-marginal of the invariant distribution of (4.2.1) is $N(0,1)$.

Proof: The proof of the above Theorem is spread over various subsections. In Section 4.2.1 we show that the process (X_t, η_t) where $\eta_t = 1/\theta_t$ is tight. This combined with the hypoelliptic condition in Section 4.2.2 shows that the process admits a invariant distribution. The marginal of the invariant distribution is identified as the target distribution in Section 4.2.3.

4.2.1 Tightness of $(X_t, \eta_t)'$

We first state and prove a lemma.

LEMMA 11. Fix $T > 0$ and an integer $k \geq 1$. Assume $E(\theta_0^{2k}) < \infty$. $\int_0^t \theta_s^k dW_s$ is a martingale with respect to $\Big\{\mathcal{F}_t = \sigma(X_s, \theta_s; 0 \leq s \leq t),\ 0 \leq t \leq T\Big\}$ and hence for any $0 \leq t \leq T$

$$E(\int_0^t \theta_s^k dW_s) \quad = \quad 0.$$

Proof: It is sufficient to show that the local martingale $Z_t := \int_0^t \theta_s^k dW_s$ is L_2-bounded for all $t \leq T$. So using Itô's isometry it suffices to show that

$$E\Big(\int_0^T \theta_s^{2k} ds\Big) < \infty.$$

Now,

$$\begin{aligned} d\theta_t &\leq q\theta_t dt \Rightarrow \theta_t \leq \theta_0 e^{qt} \\ \Rightarrow \theta_t^{2k} &\leq \theta_0^{2k} e^{2kqt} \Rightarrow E\int_0^t \theta_s^{2k} ds \leq E(\theta_0^{2k}) \frac{e^{2kqt} - 1}{2kq} < \infty, \end{aligned} \tag{4.2.2}$$

for every $t \in [0,T],\ T < \infty$. ■

4.2.1.1 Uniform boundedness of moments of X_t

We first prove a lemma that will be required in this subsection and elsewhere. Define $F_t = e^{\int_0^t \theta_u^2 du}$ and for any $k \in \mathbb{N},\ C_k := k(1-(2k-1)a)$, where $a > 0$ is a constant such that $C_k > 0$.

LEMMA 12. If $\{X_t\}$ and $\{\theta_t\}$ are solutions to (4.2.1). Fix any $k \in \mathbb{N}$ then

$$E\Big(F_t^{-C_k}\int_0^t F_u^{C_k} X_u^{2m-1}\theta_u dW_u\Big) \quad = \quad 0, \quad \text{for any } m \in \{1,2,\dots,k\}, \qquad (4.2.3)$$

where X_0 and θ_0 is such that all its moments are finite.

Proof: Fix $m \in \{1,2,\dots,k\}$. Define $\overline{F}_{t,k} := F_t^{-C_k}$. The LHS in (4.2.3) is the expectation of $Z_{t,k}(= Z_{t,k}^{(m)}) := \overline{F}_{t,k}Y_{t,k}$ where $Y_{t,k}(= Y_{t,k}^{(m)}) := \int_0^t F_u^{C_k}X_u^{2m-1}\theta_u dW_u$. We show $E(Z_{t,k}) = 0$. Applying Itô's lemma to $Z_{t,k}$ we have

$$\begin{aligned} dZ_{t,k} &= Y_{t,k}d\overline{F}_{t,k} + \overline{F}_{t,k}dY_{t,k} \\ &= -C_kY_{t,k}\theta_t^2\overline{F}_{t,k}dt + \overline{F}_{t,k}X_t^{2m-1}\theta_tF_t^{C_k}dW_t \\ &= -C_kZ_{t,k}\theta_t^2dt + X_t^{2m-1}\theta_tdW_t. \end{aligned} \qquad (4.2.4)$$

Now, taking $\tilde{Z}_{t,k} = -Z_{t,k}$, yields

$$d\tilde{Z}_{t,k} = C_kZ_{t,k}\theta_t^2dt - X_t^{2m-1}\theta_tdW_t = -C_k\tilde{Z}_{t,k}\theta_t^2dt + X_t^{2m-1}\theta_td\tilde{W}_t \qquad (4.2.5)$$

where $\tilde{W}_t = -W_t \stackrel{d}{=} W_t$. From the definition $\tilde{Z}_{0,k} = -Z_{0,k} = 0 = Z_{0,k}$. Comparing the SDE for $Z_{t,k}$ and $\tilde{Z}_{t,k}$ in (4.2.4) and (4.2.5) we see that they have the same distribution. Therefore $Z_{t,k}$ and $-Z_{t,k}$ have the same distribution, which implies that the distribution of $Z_{t,k}$ is symmetric about 0. Now to conclude $E(Z_{t,k}) = 0,\ \ \forall t \geq 0$ we show $Z_{t,k}$ has finite expectation $\forall t \geq 0$. It is sufficient to show that $E(Z_{t,k}^2) < \infty$, $\forall t \geq 0$. Now,

$$Z_{t,k}^2 = F_t^{-2C_k}\Big(\int_0^t F_s^{C_k} X_s^{2m-1}\theta_s dW_s\Big)^2 \leq \Big(\int_0^t F_s^{C_k} X_s^{2m-1}\theta_s dW_s\Big)^2$$

a.s, since $F_t^{-2C_k} \leq 1$. Therefore,

$$\begin{aligned}
E\Big(Z_{t,k}^2\Big) &\leq E\Big(\int_0^t F_s^{C_k} X_s^{2m-1}\theta_s dW_s\Big)^2 = E\Big(\int_0^t \underbrace{F_s^{2C_k}\theta_s}\,\underbrace{X_s^{4m-2}\theta_s}\, ds\Big) \\
&\leq E\Big(\Big(\int_0^t F_s^{4C_k}\theta_s^2 ds\Big)^{\frac{1}{2}}\Big(\int_0^t X_s^{8m-4}\theta_s^2 ds\Big)^{\frac{1}{2}}\Big) \\
&\leq \sqrt{E\Big(\int_0^t F_s^{4C_k}\theta_s^2 ds\Big)E\Big(\int_0^t X_s^{8m-4}\theta_s^2 ds\Big)},
\end{aligned} \tag{4.2.6}$$

where the second equality follows from Ito's Isometry and the last two inequalities follow from the Cauchy Schwartz inequality. Now for the first expectation in (4.2.6) we have

$$\begin{aligned}
E\Big(\int_0^t F_s^{4C_k}\theta_s^2 ds\Big) &= E\Big(\frac{F_t^{4C_k}-1}{4C_k}\Big) = \frac{1}{4C_k}E\Big(F_t^{4C_k}-1\Big) \\
&\leq \frac{1}{4C_k}E\Big(e^{4C_k\theta_0^2(\frac{e^{2qt}-1}{2q})}\Big) < \infty,
\end{aligned} \tag{4.2.7}$$

since from (4.2.2) $\theta_t^2 \leq \theta_0^2 e^{2qt}$. For the second term in (4.2.6) first note that from (4.2.1) and (4.2.2)

$$
\begin{aligned}
X_t &= X_0 - \int_0^t \frac{X_s\theta_s^2}{2}ds + \int_0^t \theta_s dW_s \\
\Rightarrow X_t^{8m-4}\theta_t^2 &\leq D_m^2\theta_0^2\Big(X_0^{8m-4} + \Big(\int_0^t \frac{|X_s|\theta_s^2}{2}ds\Big)^{8m-4} + \Big(\int_0^t \theta_s dW_s\Big)^{8m-4}\Big)e^{2qt} \\
\Rightarrow \int_0^t X_s^{8m-4}\theta_s^2 ds &\leq D_m^2\theta_0^2\Big(X_0^{8m-4}\int_0^s e^{2qs}ds + \int_0^t e^{2qs}\Big(\int_0^s \frac{|X_u|\theta_u^2}{2}du\Big)^{8m-4}ds \\
&+ \int_0^t e^{2qs}\Big(\int_0^s \theta_u dW_u\Big)^{8m-4}ds,
\end{aligned}
$$

for some $D_m > 0$. This implies that

$$
\begin{aligned}
\Rightarrow E\Big(\int_0^t X_s^{8m-4}\theta_s^2 ds\Big) &\leq D_m\theta_0^2\Big(E(X_0^{8m-4})\int_0^t e^{2qs}ds + E\int_0^t e^{2qs}\Big(\int_0^s \frac{|X_u|\theta_u^2}{2}du\Big)^{8m-4}ds \\
&+ E\int_0^t e^{2qs}\Big(\int_0^s \theta_u dW_u\Big)^{8m-4}ds\Big) \qquad (4.2.8)
\end{aligned}
$$

for some constant $D_m > 0$ that does not depend on X_t. Clearly the first expectation in the RHS of (4.2.8) is finite $\forall t \geq 0$.

For the second expectation in (4.2.8) we proceed as follows. From (4.2.1) we have the SDE for θ_t as

$$
\begin{aligned}
d\theta_t &= \theta_t\Big(q - \frac{|X_t|}{\sqrt{2\pi}}\theta_t\Big)dt = q\theta_t dt - \frac{|X_t|}{\sqrt{2\pi}}\theta_t^2 dt \\
\Rightarrow e^{-qt}d\theta_t - q\theta_t e^{-qt} &= -e^{-qt}\frac{|X_t|}{\sqrt{2\pi}}\theta_t^2 dt \Rightarrow d\Big(\theta_t e^{-qt}\Big) = -e^{-qt}\frac{|X_t|}{\sqrt{2\pi}}\theta_t^2 dt \\
\Rightarrow \theta_t e^{-qt} &= \theta_0 - \int_0^t e^{-qs}\frac{|X_s|}{\sqrt{2\pi}}\theta_s^2 ds \\
\Rightarrow \sqrt{\frac{\pi}{2}}\Big(\theta_0 e^{qt} - \theta_t\Big) &= \int_0^t e^{q(t-s)}\frac{|X_s|\theta_s^2}{2}ds \qquad (4.2.9)
\end{aligned}
$$

Therefore

$$\int_0^t \frac{|X_s|\theta_s^2}{2}ds \;\leq\; \int_0^t e^{q(t-s)}\frac{|X_s|\theta_s^2}{2}ds \leq \sqrt{\frac{\pi}{2}}\Big(\theta_0 e^{qt}+\theta_t\Big),$$

from (4.2.9).

Plugging the value of θ_t from (4.2.2) in (4.2.9) we have,

$$\begin{aligned}
\int_0^t \frac{|X_s|\theta_s^2}{2}ds &\leq \sqrt{2\pi}\theta_0 e^{qt} \\
\Rightarrow \Big(\int_0^t \frac{|X_s|\theta_s^2}{2}ds\Big)^{8m-4} &\leq \Big(\sqrt{2\pi}\Big)^{8m-4}\Big(\theta_0 e^{qt}\Big)^{8m-4} \\
\Rightarrow E\int_0^t e^{2qs}\Big(\int_0^s \frac{|X_u|\theta_u^2}{2}du\Big)^{8m-4}ds &\leq (2\pi)^{4m-2}\big(\int_0^t e^{(8m-2)qs}ds\big)E(\theta_0^{8m-4}) \\
&< \infty, \qquad (4.2.10)
\end{aligned}$$

for every $t \geq 0$. Hence the second expectation in the RHS of (4.2.8) is also finite $\forall t \geq 0$.

For the third term in the RHS of (4.2.8) let us define $M_s := |\int_0^s \theta_u dW_u|$ and $M_s^* = \sup_{0<u<s} M_u$. Denoting $[M]_s$ as the quadratic variation process of M_s we have $[M]_s = \int_0^s \theta_u^2 du$. Now,

$$\begin{aligned}
E(M_s)^{8m-4} &\leq E(M_s^*)^{8m-4} \leq C_m E([M_s]^{4m-2}) \\
&= C_m E\Big(\int_0^s \theta_u^2 du\Big)^{4m-2} \leq \Big(C_m \int_0^s \theta_0^2 e^{2qu}du\Big)^{4m-2}, \qquad (4.2.11)
\end{aligned}$$

where the second inequality follows from the Burkholder-Davis-Gundy (BDG) inequality and $C_m \in (0,\infty)$ is a constant. Interchanging the expectation and integrals in the third term of the RHS of (4.2.8) we get

$$\begin{aligned} E\Big(\int\limits_0^t e^{2qs}\Big(\int_0^s \theta_u dW_u\Big)^{8m-4} ds\Big) &= \int\limits_0^t e^{2qs} E\Big(\int_0^s \theta_u dW_s\Big)^{8m-4} ds \\ &= \int\limits_0^t e^{2qs} E M_s^{8m-4} ds \\ &\leq E(\theta_0^{8m-4}) \int\limits_0^t e^{2qs}\Big(C_m \int\limits_0^s e^{2qu} du\Big)^{4m-2} ds \\ &< \infty, \end{aligned} \tag{4.2.12}$$

$\forall t > 0$, where the last but one inequality follows from (4.2.11). Hence the third term of the RHS of (4.2.8) is also finite $\forall t \geq 0$. Hence combining (4.2.10) and (4.2.12) we have

$$E(Z_{t,k}^2) \quad < \quad \infty.$$

This combined with the fact that $Z_{t,k}$ is symmetric about zero proves $E(Z_{t,k}) = 0$ and hence the lemma. ■

The statement of the above lemma is true even for even powers of X, that is

LEMMA 13. Under the hypothesis of Lemma 12 the following is true

$$E\Big(F_t^{-C_k} \int\limits_0^t F_u^{C_k} X_u^{2m} \theta_u dW_u\Big) = 0 \text{ for } m \in \{0, 1, 2, \ldots, k\}.$$

Proof: We have to prove that $Z_{t,k} = \overline{F_{t,k}} Y_{t,k} := F_t^{-C_k} \int\limits_0^t F_u^{C_k} X_u^{2m} \theta_u dW_u$ has mean zero. Now

$$dZ_{t,k} \quad = \quad -C_k Z_{t,k} \theta_t^2 dt + X_t^{2m} \theta_t dW_t. \tag{4.2.13}$$

Define $\overline{Z}_{t,k} = -Z_{t,k}$ and then we see that $Z_{t,k}$ and $-Z_{t,k}$ has the same distribution. We need to show that $Z_{t,k}$ is square integrable. Following steps similar to Equation (4.2.6) of the previous lemma

$$\begin{aligned} E(Z_{t,k}^2) &\leq E\Big(\int_0^t \underbrace{F_s^{2C_k}\theta_s}\,\underbrace{X_s^{2m}\theta_s}\,ds\Big) \\ &\leq \sqrt{E\Big(\int_0^t F_s^{4C_k}\theta_s^2 ds\Big)E\Big(\int_0^t X_s^{8m}\theta_s^2 ds\Big)}. \end{aligned}$$

The first expectation is finite by Equation (4.2.7) of Lemma 4.2.3. For the second expectation we have

$$\begin{aligned} E\Big(\int_0^t X_s^{8m}\theta_s^2 ds\Big) &\leq D_m\theta_0^2\Big(E(X_0^{8m})\int_0^t e^{2qs}ds + E\int_0^t e^{2qs}\Big(\int_0^t \frac{|X_u|\theta_u^2}{2}\Big)^{8m}\Big)ds \\ &+ E\int_0^t e^{2qs}\Big(\int_0^t \theta_u dW_u\Big)^{8m} ds \end{aligned} \tag{4.2.14}$$

By applying methods in the Lemma 4.2.3 the second and the third term can be shown to be finite. This proves the lemma. ■

Here is the main lemma of this subsection.

LEMMA 14. For any $k \in \mathbb{N}$, the $2k^{th}$ ordered moment of X_t is uniformly bounded in t, i.e.,

$$\sup_{t>0} E(X_t^{2k}) < \infty,$$

if X_0 and θ_0 admit finite moments of all order.

Proof: Applying Itô's lemma to $Y_t = X_t^{2k}$ we get

$$\begin{aligned} dX_t^{2k} &= 2kX_t^{2k-1}dX_t + k(2k-1)X_t^{2k-2}\theta_t^2 dt \\ &= \Big(-kX_t^{2k}\theta_t^2 + k(2k-1)X_t^{2k-2}\theta_t^2\Big)dt + 2kX_t^{2k-1}\theta_t dW_t \\ &\leq \Big(-kX_t^{2k}\theta_t^2 + k(2k-1)(aX_t^{2k}+b)\theta_t^2\Big)dt + 2kX_t^{2k-1}\theta_t dW_t, \end{aligned}$$

since for any fixed $k \in \mathbb{N}$ and small $a > 0$, there exists $b(= b_k)$ large enough such that, $x^{2k-2} < ax^{2k} + b, \quad \forall x \in \mathbb{R}$.

Thus, for $0 < a < 1/(2k-1)$ we have

$$\begin{aligned} dX_t^{2k} &\leq -X_t^{2k}\theta_t^2\Big(k - k(2k-1)a\Big)dt \\ &+ k(2k-1)b\theta_t^2 dt + 2kX_t^{2k-1}\theta_t dW_t \\ \Rightarrow dX_t^{2k} + C_k X_t^{2k}\theta_t^2 dt &\leq k(2k-1)b\theta_t^2 dt + 2kX_t^{2k-1}\theta_t dW_t, \end{aligned} \tag{4.2.15}$$

where C_k and $F_t^{C_k}$ are defined earlier. Multiplying by the integrating factor $F_t^{C_k}$ on both sides of (4.2.15) we get

$$\begin{aligned} d\Big(X_t^{2k}F_t^{C_k}\Big) &\leq k(2k-1)b\theta_t^2 F_t^{C_k}dt + 2kF_t^{C_k}X_t^{2k-1}\theta_t dW_t \\ \Rightarrow X_t^{2k}F_t^{C_k} &\leq X_0^{2k} + k(2k-1)b\int_0^t \theta_u^2 F_u^{C_k}du + 2k\int_0^t F_u^{C_k}X_u^{2k-1}\theta_u dW_u \\ \Rightarrow X_t^{2k} &\leq X_0^{2k}F_t^{-C_k} + k(2k-1)bF_t^{-C_k}\int_0^t \theta_u^2 F_u^{C_k}du \\ &+ 2kF_t^{-C_k}\int_0^t F_u^{C_k}X_u^{2k-1}\theta_u dW_u. \end{aligned}$$

Now,

$$\int_0^t \theta_u^2 F_u^{C_k}du = (F_t^{C_k} - 1)/C_k$$

$$\begin{aligned} \Rightarrow E\Big(X_t^{2k}\Big) &\leq E\Big(F_t^{-C_k}X_0^{2k}\Big) + k(2k-1)bE\Big(\frac{1}{C_k}(1 - F_t^{-C_k})\Big) \\ &+ 2kE\Big(F_t^{-C_k}\int_0^t F_u^{C_k}X_u^{2k-1}\theta_u dW_u\Big) \end{aligned} \tag{4.2.16}$$

For the first term in (4.2.16) we have that,

$$E(F_t^{-C_k}X_0^{2k}) \leq E(X_0^{2k}) < m < \infty, \ \forall t \geq 0,$$

since $C_k\int_0^t \theta_u^2 du > 0$. Similarly $E\Big(\frac{1}{C_k}(1 - F_t^{-C_k})\Big) \leq \frac{1}{C_k}$. The third expectation is zero by Lemma 12. This proves the lemma. ■

4.2.1.2 Uniform boundedness of moments of $\eta_t = \frac{1}{\theta_t}$

LEMMA 15. *For any $k \in \mathbb{N}$, the $2k^{th}$ order moment of η_t is uniformly bounded in $t \geq 0$, i.e.,*

$$\sup_{t>0} E(\eta_t^{2k}) < \infty,$$

if X_0 and η_0 admit finite moments of all orders.

Proof. Take $\eta_t = \frac{1}{\theta_t}$. Then

$$\begin{aligned} d\eta_t &= -\frac{1}{\theta_t^2} d\theta_t \\ &= -\frac{1}{\theta_t^2}\theta_t(q - \frac{1}{\sqrt{2\pi}}|X_t|\theta_t)dt = -\eta_t(q - \frac{|X_t|}{\eta_t\sqrt{2\pi}})dt = (-\eta_t q + \frac{|X_t|}{\sqrt{2\pi}})dt \end{aligned}$$

Multiplying by the integrating factor e^{qt} on both sides of the above equation we get:

$$\begin{aligned} d(e^{qt}\eta_t) &= \frac{e^{qt}|X_t|}{\sqrt{2\pi}} dt \\ \Rightarrow e^{qt}\eta_t - \eta_0 &= \int_0^t \frac{1}{\sqrt{2\pi}} e^{qu}|X_u| du \\ \Rightarrow \eta_t &= \eta_0 e^{-qt} + \int_0^t e^{-q(t-u)}\frac{|X_u|}{\sqrt{2\pi}} du \qquad (4.2.17) \\ \Rightarrow E(\eta_t^{2k}) &= E\Big(\eta_0 e^{-qt} + \int_0^t e^{-q(t-u)}\frac{|X_u|}{\sqrt{2\pi}} du\Big)^{2k} \\ &\leq 2^{2k-1}\Big[E(\eta_0 e^{-qt})^{2k} + E\Big(\int_0^t e^{-q(t-u)}\frac{|X_u|}{\sqrt{2\pi}} du\Big)^{2k}\Big]. \qquad (4.2.18) \end{aligned}$$

Now

$$\begin{aligned} \Big(\int_0^t e^{-q(t-u)}\frac{|X_u|}{\sqrt{2\pi}} du\Big)^{2k} &= \Big(e^{-qt}\int_0^t e^{qu}\frac{|X_u|}{\sqrt{2\pi}} du\Big)^{2k} \\ &= \frac{(e^{qt}-1)^{2k}}{(qe^{qt}\sqrt{2\pi})^{2k}}\Big(\frac{q}{e^{qt}-1}\int_0^t e^{qu}|X_u| du\Big)^{2k} \\ &\leq \frac{(e^{qt}-1)^{2k}}{(q\sqrt{2\pi}e^{qt})^{2k}}\Big(\frac{q}{e^{qt}-1}\int_0^t e^{qu}|X_u|^{2k} du\Big), \end{aligned}$$

where the last inequality follows from the fact that $(E_P(|X|))^{2k} \leq E_P(|X|^{2k})$ where $k \in \mathbb{N}$ and P is any probability measure. In the above we take $P(dx) = \frac{q}{e^{qt}-1}e^{qx}dx$ on

$[0, t]$. Therefore interchanging the expectation and integrals on the last term of 4.2.17 we have

$$
\begin{aligned}
E(\eta_t^{2k}) &\leq 2^{2k-1}\Big[E(\eta_0^{2k})e^{-2kqt} + \frac{(e^{qt}-1)^{2k}}{(q\sqrt{2\pi}e^{qt})^{2k}}\frac{q}{e^{qt}-1}\int_0^t e^{qu}E(|X_u|^{2k})du\Big] \\
&\leq 2^{2k-1}\Big[E(\eta_0^{2k}) + \frac{(e^{qt}-1)^{2k}}{(q\sqrt{2\pi}e^{qt})^{2k}}M_0\Big] \\
&\leq M_1 < \infty \qquad (4.2.19)
\end{aligned}
$$

where the last but one inequality follows from Lemma 4.2.1.1 that even moments of X_t are uniformly bounded in $t \geq 0$. ■

Remark 16. From (4.2.19) it is evident that for all $t > 0$, there is a null set, outside of which $\theta_t = \frac{1}{\eta_t} > 0$ whenever $\theta_0 > 0$, as otherwise, $\sup\limits_{t>0} E(\eta_t^{2k})$ would be infinity. Again, from the proof above, it is clear that

$$
\eta_t = \eta_0 e^{-qt} + \int_0^t e^{-q(t-u)}\frac{|X_u|}{\sqrt{2\pi}}du > 0, \quad \text{whenever } \eta_0 \geq 0.
$$

Combining the above two lemmas we have the following tightness result for the vector $(X_t, \eta_t)'$.

4.2.1.3 Tightness

Lemma 16. If X_0 and θ_0 admits moments of all orders and $\theta_0 > 0$ $a.s.$then, for the coupled system (4.2.1) joint distribution of $\{(X_t, \eta_t)' : t \geq 0\}$ is tight.

Proof. Let R_1 and R_2 be two positive numbers. Then

$$
\begin{aligned}
P(|X_t| < R_1, |\eta_t| < R_2) &= 1 - P((|X_t| > R_1) \cup (|\eta_t| > R_2)) \\
&> 1 - (P(|X_t| > R_1) + P(|\eta_t| > R_2)) \\
&> 1 - E(|X_t|)/R_1 - E(|\eta_t|)/R_2.
\end{aligned}
$$

Hence given any $\epsilon > 0$ we can choose R_1, R_2 sufficiently large so that $P(|X_t| < R_1, |\eta_t| < R_2) > 1 - \epsilon$. This proves the tightness of $(X_t, \eta_t)'$. ■

4.2.1.4 Finiteness of Time average of moments of θ_t

In this section C will stand for a generic finite constant that might take different values in different situations. We assume throughout that X_0 and θ_0 admit finite moments of all orders. For non-random initial data this is trivially true.

LEMMA 17. Let X_0 and θ_0 admit finite moments of all order. Then

$$\sup_{t>1} \frac{1}{t} \int_0^t E(\theta_u^{\frac{k}{2}})du \quad < \quad C \text{ for every } k \in \mathbb{N}.$$

Proof: We proceed sequentially through the following steps.

Step 1: We first prove

$$\sup_{t>1} \frac{1}{t} \int_0^t E(|X_u|\theta_u)du < \infty.$$

This fact will be used in Step 2. To prove this note that

$$\begin{aligned}
d(1+\theta_t) &= d\theta_t = \theta_t(q - |X_t|\theta_t/\sqrt{2\pi})dt \\
&= q\theta_t dt - \frac{(1+\theta_t)|X_t|\theta_t}{\sqrt{2\pi}}dt + \frac{|X_t|\theta_t}{\sqrt{2\pi}}dt \\
\Rightarrow d(1+\theta_t) + \frac{(1+\theta_t)|X_t|\theta_t}{\sqrt{2\pi}}dt &= q\theta_t dt + \frac{|X_t|\theta_t}{\sqrt{2\pi}}dt \\
\Rightarrow \frac{d(1+\theta_t)}{1+\theta_t} + \frac{1}{\sqrt{2\pi}}|X_t|\theta_t dt &= \frac{\theta_t}{1+\theta_t}\Big(q + \frac{|X_t|}{\sqrt{2\pi}}\Big)dt \\
&\leq \Big(q + \frac{|X_t|}{\sqrt{2\pi}}\Big)dt \\
\Rightarrow \log\frac{1+\theta_t}{1+\theta_0} + \frac{1}{\sqrt{2\pi}}\int_0^t |X_u|\theta_u du &\leq qt + \frac{1}{\sqrt{2\pi}}\int_0^t |X_u|du \\
\Rightarrow \frac{1}{t}\int_0^t |X_u|\theta_u du &\leq \sqrt{2\pi}q + \frac{1}{t}\int_0^t |X_u|du \\
&+ \sqrt{2\pi}\frac{\log(1+\theta_0)}{t}. \qquad (4.2.20)
\end{aligned}$$

Thus, $\frac{1}{t}\int\limits_0^t E(|X_u|\theta_u)du \leq \sqrt{2\pi}q + \frac{1}{t}\int\limits_0^t E(|X_u|)du + \sqrt{2\pi}\frac{E(\log(1+\theta_0))}{t}$. Therefore, using the moment bounds for X_t from Section 4.2.1.1,

$$\sup_{t>1}\frac{1}{t}\int\limits_0^t E(|X_u|\theta_u)du \quad < \quad C. \tag{4.2.21}$$

<u>Step 2:</u> We now prove by induction, that for any $k \in \mathbb{N}$,

$$\sup_{t>1}\frac{1}{t}\int\limits_0^t E(\theta_u^{\frac{k}{2}})du < C. \tag{4.2.22}$$

Let, as before, $\eta_t = \frac{1}{\theta_t}$ then $d\eta_t = (-q\eta_t + |X_u|/\sqrt{2\pi})dt$.

Applying Itô's lemma to $Y_t = X_t^2\eta_t^{2-k/2}$, with $k \in \mathbb{N}$, we get

$$\begin{aligned}
dY_t &= 2X_t\eta_t^{2-k/2}dX_t + (2-k/2)X_t^2\eta_t^{1-k/2}d\eta_t + \frac{1}{2}2\eta_t^{2-k/2}(dX_t)^2 \\
&= 2X_t\eta_t^{2-k/2}(-\frac{X_t}{2\eta_t^2}dt + \frac{1}{\eta_t}dW_t) + (2-k/2)X_t^2\eta_t^{1-k/2}(-q\eta_t dt + \frac{|X_t|}{\sqrt{2\pi}}dt) \\
&+ \eta_t^{2-k/2}\eta_t^{-2}dt \\
&= \Big(-X_t^2\eta_t^{-k/2} - q(2-k/2)X_t^2\eta_t^{2-k/2} + \frac{2-k/2}{\sqrt{2\pi}}|X_t|^3\eta_t^{1-k/2} + \eta_t^{-k/2}\Big)dt \\
&+ 2X_t\eta_t^{1-k/2}dW_t.
\end{aligned} \tag{4.2.23}$$

Thus, integrating both sides from 0 to t, rearranging and dividing by t and then taking expectations we get

$$\begin{aligned}
\int_0^t \theta_s^{\frac{k}{2}} ds &= X_t^2 \eta_t^{\frac{4-k}{2}} - X_0^2 \eta_0^{\frac{4-k}{2}} + \int_0^t X_s^2 \theta_s^{\frac{k}{2}} ds \\
&+ \frac{(4-k)q}{2} \int_0^t X_s^2 \eta_s^{\frac{4-k}{2}} ds - \frac{2-k/2}{\sqrt{2\pi}} \int_0^t |X_s|^3 \eta_s^{\frac{2-k}{2}} ds \\
&- 2 \int_0^t X_s \eta_s^{\frac{2-k}{2}} dW_s \\
\Rightarrow \frac{1}{t} \int_0^t \theta_s^{\frac{k}{2}} ds &= \frac{1}{t}(X_t^2 \eta_t^{\frac{4-k}{2}} - X_0^2 \eta_0^{\frac{4-k}{2}}) + \frac{1}{t} \int_0^t X_s^2 \theta_s^{\frac{k}{2}} ds \\
&+ \frac{(4-k)q}{2t} \int_0^t X_s^2 \eta_s^{\frac{4-k}{2}} ds \\
& -\frac{2-k/2}{t\sqrt{2\pi}} \int_0^t |X_s|^3 \eta_s^{\frac{2-k}{2}} ds - \frac{2}{t} \int_0^t X_s \eta_s^{\frac{2-k}{2}} dW_s \\
\Rightarrow \frac{1}{t} \int_0^t E(\theta_s^{\frac{k}{2}}) ds &= \frac{1}{t} E(X_t^2 \eta_t^{\frac{4-k}{2}} - X_0^2 \eta_0^{\frac{4-k}{2}}) + \frac{1}{t} \int_0^t E(X_s^2 \theta_s^{\frac{k}{2}}) ds \\
&+ \frac{(4-k)q}{2} \frac{1}{t} \int_0^t E(X_s^2 \eta_s^{\frac{4-k}{2}}) ds \\
&+ \frac{k-4}{2\sqrt{2\pi}} \frac{1}{t} \int_0^t E(|X_s|^3 \eta_s^{\frac{2-k}{2}}) ds \\
&+ 2\frac{1}{t} E\Big(-\int_0^t (X_s \eta_s^{\frac{2-k}{2}}) dW_s\Big). \qquad (4.2.24)
\end{aligned}$$

Now for any $k \in \mathbb{N}$ we have,

$$\begin{aligned}\frac{1}{t}\int_0^t X_s^2\theta_s^{\frac{k}{2}}ds &= \frac{1}{t}\int_0^t (|X_s|^{\frac{k}{k+1}}\theta_s^{\frac{k}{2}})(|X_s|^{\frac{k+2}{k+1}})ds \\ &\leq \Big(\frac{1}{t}\int_0^t |X_s|\theta_s^{\frac{k+1}{2}}ds\Big)^{\frac{k}{k+1}}\Big(\frac{1}{t}\int_0^t |X_s|^{k+2}ds\Big)^{\frac{1}{k+1}}, \qquad (4.2.25)\end{aligned}$$

which follows from the Holder's inequality with $p = \frac{k+1}{k}$ and $q = k+1$. Therefore,

$$\begin{aligned}E\Big(\frac{1}{t}\int_0^t X_s^2\theta_s^{\frac{k}{2}}\Big) &\leq E\Big(\Big(\frac{1}{t}\int_0^t |X_s|\theta_s^{\frac{k+1}{2}}ds\Big)^{\frac{k}{k+1}}\Big(\frac{1}{t}\int_0^t |X_s|^{k+2}ds\Big)^{\frac{1}{k+1}}\Big) \\ &\leq \Big(E\Big(\frac{1}{t}\int_0^t |X_s|\theta_s^{\frac{k+1}{2}}ds\Big)\Big)^{\frac{k}{k+1}} \times \Big(E\Big(\frac{1}{t}\int_0^t |X_s|^{k+2}ds\Big)\Big)^{\frac{1}{k+1}} \\ &= \Big(\frac{1}{t}\int_0^t E(|X_s|\theta_s^{\frac{k+1}{2}})ds\Big)^{\frac{k}{k+1}}\Big(\frac{1}{t}\int_0^t E(|X_s|^{k+2})ds\Big)^{\frac{1}{k+1}}, \qquad (4.2.26)\end{aligned}$$

where the last inequality follows from Holder's inequality with $p = \frac{k+1}{k}$ and $q = k+1$. Therefore,

$$\begin{aligned}\frac{1}{t}\int_0^t E(X_s^2\theta_s^{\frac{k}{2}})ds &\leq \Big(\frac{1}{t}\int_0^t E(|X_s|\theta_s^{\frac{k+1}{2}})ds\Big)^{\frac{k}{k+1}} \\ &\times \Big(\frac{1}{t}\int_0^t E(|X_s|^{k+2})ds\Big)^{\frac{1}{k+1}}. \qquad (4.2.27)\end{aligned}$$

Again $\forall k \geq 2$

$$
\begin{aligned}
d\theta_t^{\frac{k-1}{2}} = \frac{k-1}{2}\theta_t^{\frac{k-1}{2}-1}d\theta_t &= \frac{k-1}{2}\theta_t^{\frac{k-1}{2}}(q-\frac{|X_t|\theta_t}{\sqrt{2\pi}})dt \\
&= \frac{q(k-1)}{2}\theta_t^{\frac{k-1}{2}}dt - \frac{k-1}{2}\frac{|X_t|\theta_t^{\frac{k+1}{2}}}{\sqrt{2\pi}}dt \\
\Rightarrow \frac{1}{\sqrt{2\pi}}\int_0^t |X_s|\theta_s^{\frac{k+1}{2}}ds &= q\int_0^t \theta_s^{\frac{k-1}{2}}ds - \frac{2}{k-1}(\theta_t^{\frac{k-1}{2}} - \theta_0^{\frac{k-1}{2}}) \\
\Rightarrow \frac{1}{t}\int_0^t E\Big(|X_s|\theta_s^{\frac{k+1}{2}}ds\Big) &\leq \sqrt{2\pi}q\frac{1}{t}\int_0^t E\Big(\theta_s^{\frac{k-1}{2}}ds\Big) + \frac{2\sqrt{2\pi}}{k-1}\frac{1}{t}E\Big(\theta_0^{\frac{k-1}{2}}\Big) \\
&. \qquad (4.2.28)
\end{aligned}
$$

Plugging (4.2.28) in (4.2.27)

$$
\begin{aligned}
\frac{1}{t}\int_0^t E(X_s^2\theta_s^{\frac{k}{2}})ds &\leq \Big(\sqrt{2\pi}q\frac{1}{t}\int_0^t E\Big(\theta_s^{\frac{k-1}{2}}\Big)ds + \frac{2\sqrt{2\pi}}{k-1}\frac{1}{t}E\Big(\theta_0^{\frac{k-1}{2}}\Big)\Big)^{\frac{k}{k+1}} \\
&\times \Big(\frac{1}{t}\int_0^t E(|X_s|^{k+2})ds\Big)^{\frac{1}{k+1}}. \qquad (4.2.29)
\end{aligned}
$$

And finally plugging (4.2.29) in (4.2.24) we get for $k \geq 2$

$$
\begin{aligned}
\frac{1}{t}\int_0^t E(\theta_s^{\frac{k}{2}})ds \;\; &\leq \;\; \frac{1}{t}E(X_t^2\eta_t^{\frac{4-k}{2}} - X_0^2\eta_0^{\frac{4-k}{2}}) \\
&+ \;\; \Big(\sqrt{2\pi}q\frac{1}{t}\int_0^t E\Big(\theta_s^{\frac{k-1}{2}}\Big)ds + \frac{2\sqrt{2\pi}}{k-1}\frac{1}{t}E(\theta_0^{\frac{k-1}{2}})\Big)^{\frac{k}{k+1}} \\
&\times \;\; \Big(\frac{1}{t}\int_0^t E(|X_s|^{k+2})ds\Big)^{\frac{1}{k+1}} \\
&+ \;\; \frac{(4-k)q}{2}\frac{1}{t}\int_0^t E(X_s^2\eta_s^{\frac{4-k}{2}})ds \\
&+ \;\; \frac{k-4}{2\sqrt{2\pi}}\frac{1}{t}\int_0^t E(|X_s|^3\eta_s^{\frac{2-k}{2}})ds \\
&+ \;\; 2\frac{1}{t}E\Big(-\int_0^t (X_s\eta_s^{\frac{2-k}{2}})dW_s\Big). \qquad (4.2.30)
\end{aligned}
$$

To prove $\sup_{t>1}\frac{1}{t}\int_0^t E(\theta_s^{\frac{k}{2}})ds$ is finite $\forall\; k \in \mathbb{N}$ we proceed by induction:

Step 2a: For k=1 we consider Equation (4.2.24). By an application of the Young's inequality and the fact that all the moments of X_s and η_s are uniformly bounded (proved earlier in Lemma 14 and 15) we have:

$$
\begin{aligned}
\sup_{t>1}\frac{1}{t}E(X_t^2\eta_t^{\frac{4-1}{2}} - X_0^2\eta_0^{\frac{4-1}{2}}) \;\; &< \;\; C, \\
\sup_{t>1}\frac{1}{t}\int_0^t E(X_s^2\eta_s^{\frac{4-1}{2}})ds \;\; &< \;\; C, \\
\text{and} \\
\sup_{t>1}\frac{1}{t}\int_0^t E(|X_s|^3\eta_s^{\frac{2-1}{2}})ds \;\; &< \;\; C.
\end{aligned}
$$

This proves that the supremum over $t > 1$ of the first, third and fourth term in the RHS of (4.2.24) is finite. The supremum of the second term of (4.2.24) is bounded by the RHS of (4.2.27), whose first term is finite by (4.2.21) of Step 1 and the second term is finite by the uniform boundedness of moments of X. The fourth term in the RHS of (4.2.24) is negative. Therefore we are left with only the Itô integral or the last term of (4.2.24). Now,

$$E\Big(\int_0^t X_s \eta_s^{\frac{1}{2}} dW_s\Big)^2 = E\int_0^t X_s^2 \eta_s ds$$

is finite $\forall t \geq 0$ by an application of Young's inequality and the uniform boundedness of all the moments of X_t and η_t. Therefore $\int_0^t X_s \eta_s^{\frac{1}{2}} dW_s$ is a square integrable martingale and hence

$$\sup_{t>1} \frac{1}{t} E\Big(-\int_0^t X_s \eta_s^{\frac{1}{2}} dW_s\Big) = 0.$$

This completes the proof that $\sup_{t>1} \frac{1}{t}\int_0^t E(\theta_s^{\frac{1}{2}})ds$ is finite.

Step 2b: Assume that the hypothesis is true for $k \leq m-1$, for $m \geq 2$ i.e.,

$$\sup_{t>1} \frac{1}{t}\int_0^t E(\theta_s^{\frac{k}{2}})ds \quad < \quad C, \quad k \leq m-1.$$

Step 2c: Consider $k = m \geq 2$. In this case we consider Equation (4.2.30).
For $m = 2$ the fourth in RHS of (4.2.30) is negative. Supremum of the other terms is finite by the moment bounds of X_s and η_s and by the proof that $\sup_{t>1}\int_0^t E(\theta_s^{\frac{1}{2}})ds < \infty$ in Step 2a.

For $m = 3, 4$ the supremum of the first term in the RHS of (4.2.30) is finite (by the arguments given in 2a). The supremum of the second (product) term is finite by the induction hypothesis (in 2b) and by the finiteness of the moments of X_s. The supremum of the third term is finite by the finiteness of the moments of X_t and η_t. The fourth term is negative for $m = 3$ or zero for $m = 4$. Hence it is bounded by zero.

For the fifth (Itô Integral) term in (4.2.30) we first apply the Itô's lemma and then Cauchy Schwartz inequality to get

$$\begin{aligned}\int_0^t E(X_s^2\theta_s^{m-2})ds \leq \int_0^t \sqrt{E(X_s^4)E(\theta_s^{2(m-2)})}ds &\leq C\int_0^t \sqrt{E(\theta_s^{2(m-2)})}ds \\ &< \infty, \end{aligned} \tag{4.2.31}$$

since θ_t is bounded as in Equation (4.2.2). Thus, $\int_0^t X_s\theta_s^{\frac{m-2}{2}}dW_s$ is a square integrable martingale with respect to the given filtration over any finite interval $[0,T]$ and therefore the expectation is zero. Consequently,

$$\sup_{t>1}\frac{1}{t}E\Big(-\int_0^t X_s\eta_s^{\frac{2-k}{2}}dW_s\Big) = \sup_{t>1}\frac{1}{t}E\Big(-\int_0^t X_s\theta_s^{\frac{k-2}{2}}dW_s\Big) = 0.$$

Next consider $m > 4$. Then third term in the RHS of (4.2.30) is negative. For the first term in the RHS of (4.2.30) apply Young's inequality with $p = m-3$ and $q = \frac{m-3}{m-4}$ to get

$$\begin{aligned} X_s^2\theta_s^{\frac{m-4}{2}} \leq \frac{1}{m-3}X_s^{2(m-3)} + \frac{m-4}{m-3}\theta_s^{\frac{m-3}{2}} \quad &\Rightarrow \quad E(X_s^2\theta_s^{\frac{m-4}{2}}) \leq \frac{1}{m-3}E(X_s^{2(m-3)}) \\ &+ \quad \frac{m-4}{m-3}E(\theta_s^{\frac{m-3}{2}}) \\ \Rightarrow \sup_{t>1}\frac{1}{t}\int_0^t E(X_s^2\theta_s^{\frac{m-4}{2}})ds \quad &\leq \quad \frac{1}{m-3}\sup_{t>1}\frac{1}{t}\int_0^t E(X_s^{2(m-3)})ds \\ &+ \quad \frac{m-4}{m-3}\sup_{t>1}\frac{1}{t}\int_0^t E(\theta_s^{\frac{m-3}{2}})ds \\ &< \quad \infty, \end{aligned}$$

which follows from the fact that moments of X_t are uniformly bounded and the second term is finite by the induction hypothesis. Consequently, the supremum of the first term in the RHS of (4.2.30) is finite.

The supremum of the second (product) term is finite by the induction hypothesis and by the finiteness of the moments of X_s (as argued in the case $m = 3, 4$ above).

The fourth term we apply the Young's inequality with $p = m-1$ and $q = \frac{m-1}{m-2}$ to get:

$$
\begin{aligned}
|X_s|^3\theta_s^{\frac{m-2}{2}} \leq \frac{|X_s|^{3p}}{p} + \frac{\theta_s^{\frac{q(m-2)}{2}}}{q} &= \frac{1}{m-1}|X_s|^{3(m-1)} + \frac{m-2}{m-1}\theta_s^{\frac{m-1}{2}}, \\
\Rightarrow \sup_{t>1}\frac{1}{t}\int_0^t E\Big(|X_s|^3\theta_s^{\frac{m-2}{2}}\Big)ds &\leq \frac{1}{m-1}\sup_{t>1}\frac{1}{t}\int_0^t E\Big(|X_s|^{3(m-1)}\Big)ds \\
&\quad + \frac{m-2}{m-1}\sup_{t>1}\frac{1}{t}\int_0^t E\Big(|\theta_s|^{\frac{m-1}{2}}\Big)ds \\
&< \infty, \qquad (4.2.32)
\end{aligned}
$$

which follows from the fact that the moments of X_t are uniformly bounded in t and by the induction hypothesis.

For the fifth term we argue as in (4.2.31) to infer that it is a square integrable martingale with respect to the given filtration over any finite interval $[0, T]$ and hence the expectation is zero.

Therefore the supremum of the LHS of (4.2.30) is finite for all $m \geq 2$. Thus the Steps 2a, 2b and 2c complete the proof of Step 2 (4.2.22) and therefore Lemma 17 is proved. ■

4.2.2 Hypoelliptic condition

Here we show that the vector fields corresponding to (4.2.1) satisfies the Hörmander's hypoelliptic condition (see Proposition 15 below for the statement of the condition). Since the condition requires smooth vector fields, we convert the drift and diffusion coefficients in (3.4.1) into smooth vector fields.

For this purpose, define

$$
b_\epsilon(x,\eta) = \left(-\frac{x}{2\eta^2},\ -q\eta + \frac{g_\epsilon(x)}{\sqrt{2\pi}}\right), \qquad (4.2.33)
$$

where $g_\epsilon(x)$, a smooth function $\to |x|$ as $\epsilon \downarrow 0$ in the point-wise limits and $\sigma(x,\eta) = \begin{pmatrix} 1/\eta & 0 \\ 0 & 0 \end{pmatrix}$ as the drift and the diffusion coefficient respectively of the equation with the re-parametrisation $\eta = \frac{1}{\theta}$. Such function $g_\epsilon(\cdot)$ can be constructed by convolving the

function $|x|$ with a mollifier (for example $\frac{1}{\sqrt{2\pi\epsilon}}e^{-\frac{1}{2\epsilon^2}x^2}$).
Consider an SDE in the Stratonovich form:

$$dX_t = A_0(X_t)dt + \sum_{\alpha=1}^{n} A_\alpha(X_t) \circ dW_t^\alpha. \tag{4.2.34}$$

where $A_0, \{A_\alpha : \alpha = 1, \dots, n\}$ is a smooth vector fields on a differential manifold M and $\circ$ denotes Stratonovich integral. The SDE in the Itô form and the Stratonovich form are interchangeable. For a multidimensional SDE, given in the Itô's form,

$$d\mathbf{X}_t = \mathbf{b}(t, \mathbf{X}_t)dt + \sigma(t, \mathbf{X}_t)d\mathbf{W}_t$$

can be readily converted into the Stratonovich form from the following equation:

$$\tilde{b}_i(t, \mathbf{x}) = b_i(t, \mathbf{x}) - \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{n} \frac{\partial \sigma_{i,j}}{\partial x_k}\sigma_{k,j}; \quad 1 \le i \le n$$

where $\tilde{\mathbf{b}}(t, x) = (\tilde{b}_i(t, x))'$ is the drift term for the Stratonovich form. In our case, $p = n = 2$ and from the form of σ in (4.2.1), we find that $\tilde{b^\epsilon}$ and b^ϵ are the same and it equals A_0. We identify the diffusion coefficients $A_1(\mathbf{X}_t) = (\eta, 0)'$ and $A_2(\mathbf{X}_t) = (0, 0)'$ as vector fields in M, here upper half plane of $\mathbb{R}^2$. Here is the condition due to Hörmander [34]:

Proposition 15. Let $\{A_0, A_1, \dots, A_n\}$ be $n+1$ smooth vector fields on a smooth manifold M. Define the Lie Bracket $[V, W]$ between two vector fields V and W as another vector field on M defined in the following manner

$$[V, W](f) = V(W(f)) - W(V(f)) \ \ \forall f \in C^\infty(M).$$

The Hörmander's hypoelliptic condition is satisfied if :

$$\begin{gathered} A_{j_0}(\mathbf{y}),\ [A_{j_0}(\mathbf{y}), A_{j_1}(\mathbf{y})],\ [[A_{j_0}(\mathbf{y}), A_{j_1}(\mathbf{y})], A_{j_2}(\mathbf{y})], \\ \dots\ [[[[A_{j_0}(\mathbf{y}), A_{j_1}(\mathbf{y})], A_{j_2}(\mathbf{y})], A_{j_3}(\mathbf{y})], \dots, A_{j_k}(\mathbf{y})] \end{gathered}$$

spans M for every $\mathbf{y} \in M$ and any $1 \le j_0 \le n$ and $\{j_1, \dots, j_k\} \in \{0, 1, \dots, n\},\ k \ge 1$.

Lemma 18. The vector fields $A_0^\epsilon(\mathbf{y})$ and $A_1(\mathbf{y})$ satisfy Hörmander's hypoelliptic condition of Proposition 15.

Proof: Identifying (4.2.34) with (4.2.1) we have (writing $\mathbf{y} = (x, \eta)'$):

$$\begin{aligned} A_0^\epsilon(\mathbf{y}) &= -\frac{x}{2\eta^2}\frac{\partial}{\partial x} + (-q\eta + \frac{g_\epsilon(x)}{\sqrt{2\pi}})\frac{\partial}{\partial \eta}, \\ A_1(\mathbf{y}) &= \frac{1}{\eta}\frac{\partial}{\partial x}. \end{aligned}$$

Therefore the vectors corresponding to $A_1(\mathbf{y})$ and $[A_1(\mathbf{y}), A_0^\epsilon(\mathbf{y})]$ will be $\left(\frac{1}{\eta}, 0\right)^T$ and $\left(\frac{1}{\eta^2}\left(-\frac{1}{2\eta} - q\eta + \frac{1}{\sqrt{2\pi}}g_\epsilon(x)\right), \frac{1}{\sqrt{2\pi}}\frac{1}{\eta}g_\epsilon'(x)\right)^T$. Note, $\theta_t = 1/\eta_t > 0$ almost surely, since by Lemma 15 we have $\sup_{t>0} E(\eta_t^2) < \infty$. Thus, the zero set of $\{X_t\}$ has Lebesgue measure zero almost surely since the zero set of $\{W_t\}$ has Lebesgue measure zero. Therefore these two vector fields span the upper half plane of $\mathbb{R}^2$, for $x \neq 0$. Also, for $x \neq 0$, we can take $\epsilon \to 0$ and get the same result. Note that the convergence is uniform over each compacts in the set $\{(x, \eta) : x \neq 0, \eta > 0\}$. ∎

Remark 17. In the case of the Normal mollifier i.e,

$$\begin{aligned} g_\epsilon(y) &= \frac{1}{\sqrt{2\pi}\epsilon}\int_{-\infty}^{\infty} |x| e^{-\frac{1}{2\epsilon^2}(y-x)^2} dx \\ &= \frac{1}{\sqrt{2\pi}\epsilon}\int_0^{\infty} x e^{-\frac{1}{2\epsilon^2}(y-x)^2} dx + \frac{1}{\sqrt{2\pi}\epsilon}\int_{-\infty}^{0} (-x) e^{-\frac{1}{2\epsilon^2}(y-x)^2} dx. \end{aligned}$$

For the first integral

$$\begin{aligned} \frac{1}{\sqrt{2\pi}\epsilon}\int_0^{\infty} x e^{-\frac{1}{2\epsilon^2}(y-x)^2} dx &= \frac{1}{\sqrt{2\pi}}\int_{-\frac{y}{\epsilon}}^{\infty} (y + \epsilon z) e^{-\frac{z^2}{2}} dz, \text{ substituting } z = \tfrac{x-y}{\epsilon}, \\ &= y(1 - \Phi(-\frac{y}{\epsilon})) + \frac{1}{\sqrt{2\pi}}\epsilon \int_{\frac{y^2}{2\epsilon^2}}^{\infty} e^{-t} dt, \text{ substituting } t = \tfrac{z^2}{2}, \\ &= y\Phi(\frac{y}{\epsilon}) + \frac{1}{\sqrt{2\pi}}\epsilon e^{-\frac{y^2}{2\epsilon^2}}, \end{aligned}$$

where $\Phi(\cdot)$ is the distribution function of the standard Normal variable. Similarly for the second integral we have

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}\epsilon}\int_{-\infty}^{0}(-x)e^{-\frac{1}{2\epsilon^2}(y-x)^2}dx &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{y}{\epsilon}}-\overline{y+\epsilon z}e^{-\frac{z^2}{2}}dz, \text{ substituting } z=\tfrac{x-y}{\epsilon}, \\
&= -y\Phi(-\frac{y}{\epsilon})-\frac{1}{\sqrt{2\pi}}\epsilon\int_{-\infty}^{-\frac{y}{\epsilon}} ze^{-\frac{z^2}{2}}dz \\
&= -y\Phi(-\frac{y}{\epsilon})-\epsilon\frac{1}{\sqrt{2\pi}}\int_{\infty}^{\frac{y^2}{2\epsilon^2}} e^{-t}dt \\
&= -y\Phi(-\frac{y}{\epsilon})+\epsilon\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2\epsilon^2}} \\
\Rightarrow g_\epsilon(y) &= y\Big(\Phi(\frac{y}{\epsilon})-\Phi(-\frac{y}{\epsilon})\Big)+2\epsilon\phi(\frac{y}{\epsilon}) \\
\Rightarrow \frac{d}{dy}g'_\epsilon(y) &= \Phi(\frac{y}{\epsilon})-\Phi(-\frac{y}{\epsilon})+2\frac{y}{\epsilon}\phi(\frac{y}{\epsilon})-2\frac{y}{\epsilon}\phi(\frac{y}{\epsilon}) \\
\Rightarrow |\frac{d}{dy}g_\epsilon(y)| &\leq |\Phi(\frac{y}{\epsilon})-\Phi(-\frac{y}{\epsilon})|,
\end{aligned}$$

where $\phi(\cdot)$ is the density function of the standard Normal variable. Now for any $\epsilon > 0$ and any $y \in \mathbb{R}$ we have

$$|\Phi(\frac{y}{\epsilon})-\Phi(-\frac{y}{\epsilon})| \leq 1$$

which implies that

$$\begin{aligned}
|\frac{d}{dy}g_\epsilon(y)| &\leq 1 \ \forall\epsilon>0 \\
\Rightarrow \sup_\epsilon|\frac{d}{dy}g_\epsilon(y)| &< \infty,\ \forall\ y\in\mathbb{R},
\end{aligned}$$

which implies that the family $\{g_\epsilon(\cdot)\}$ is equicontinuous.

It is well known that if the vector fields $A_0(\mathbf{y})$ and $A_1(\mathbf{y})$ satisfy the above conditions then the solution of the SDE (4.2.34) admits a smooth transition density (see, for example Nualart [49]).

Hence, even though the original diffusion is singular its transition probability has density (see Kliemann [38]). Again, since the coupled diffusion is tight, it admits unique invariant probability by Kliemann [38] which admits a density.

REMARK 18. Note that although we are interested in the distribution of $\{X_t\}$ showing tightness of the process $\{X_t\}$ only it would not suffice since θ_t may be a function of $\{X_s; 0 \leq s \leq t\}$, so marginally $\{X_t\}$ may not be a Markov process. Hence $\sup_{t>0} E|X_t| < M$ would give the tightness of X but it would not be possible to say anything about the existence of a unique invariant distribution of $\{X_t\}$.

REMARK 19. Consider a 1 dimensional process $\{X_t\}$ which satisfies the SDE

$$dX_t = b(X_t)dt + a(X_t)dW_t.$$

The invariant distribution $\pi(\cdot)$ of X_t is given by solving

$$L^*\pi(x) \quad = \quad 0,$$

which gives

$$\pi(x) \quad = \quad \frac{c}{a(x)} exp(\int_0^x \frac{2b(u)}{a(u)} du).$$

Here, L^* is the adjoint of the infinitesimal operator given by

$$L^*\pi(x) \quad = \quad -\frac{\partial}{\partial x}(b(x)\pi(x)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(a(x)\pi(x)).$$

Now suppose we have a sequence of processes $\{X_{n,t}\}$ whose drift function is given by $\{b_n(x)\}$ for suitable functions $b_n(\cdot)$ then the respective invariant distributions will be

$$\pi_n(x) \quad = \quad \frac{c}{a(x)} exp(\int_0^x \frac{2b_n(u)}{a(u)} du).$$

Now if $b_n(\cdot)$ are equicontinuous and uniformly bounded in a compact support and $a(\cdot)$ is bounded away from zero then $\pi(\cdot)$ will also be equicontinuous and uniformly bounded in a compact support. We try to extend this for a two dimensional process (X_t, η_t) in Remark 21.

REMARK 20. Consider the drift term given in Equation (4.2.33). Take $\epsilon = \frac{1}{n}$ for some $n \in \mathbb{N}$ and let $(X_{n,t}, \eta_{n,t})$ be the corresponding process. By the hypoelliptic condition the limiting distribution of $(X_{n,t}, \eta_{n,t})$ exists and has a density which we denote as $\pi_n(x, \eta)$. Now we have already proved that the moments of $|X_t|$ is uniformly bounded over all $t > 0$. Therefore from weak convergence and uniform integrability we have that the moments the limiting distribution of X_t is bounded. Further, since the standard Normal density is of the form given in Equation (5.2.2) of Chapter 5, from Theorem 7 and Remark 28 of the same chapter we can say that $\limsup_{t\to\infty} E(e^{sX_t}) < \infty, \forall s \in \mathbb{R}$ and $\limsup_{t\to\infty} E(e^{s\eta_t}) < \infty$, $\forall s \in \mathbb{R}$. Since moment bound of $\{X_{n,t}, \eta_{n,t}\}$ are similar to that of $\{X_t, \eta_t\}$ uniformly in n, their m.g.fs would also have the same property uniformly in n. Consequently, by Cauchy-Schwarz inequality, their joint m.g.f is also finite uniformly in n for all $s_1, s_2 \in \mathbb{R}$. Therefore, joint m.g.f of the limiting distribution of $\{X_{n,t}, \eta_{n,t}\}$, i.e. the m.g.f of $\pi_n(\cdot, \cdot)$ is finite for all $s_1, s_2 \in \mathbb{R}$ uniformly in n.

Remark 21. Consider the SDE whose drift coefficient $b_n(\cdot, \cdot)$ is defined as in Equation 4.2.33 with $\epsilon = \frac{1}{n}$ for some $n \in \mathbb{N}$. Let L_n denote the infinitesimal generator of the process $(X_{n,t}, \eta_{n,t})$, i.e.,

$$L_n f = (-\frac{x}{2\eta^2})\frac{\partial}{\partial x} f + (-q\eta + \frac{g_n(x)}{\sqrt{2\pi}})\frac{\partial}{\partial \eta} f + \frac{1}{2\eta^2}\frac{\partial^2}{\partial x^2} f,$$

where $g_n(\cdot) = g_{\frac{1}{n}}(\cdot)$ and $f(\cdot, \cdot)$ belongs to the domain of L_n. Then the adjoint of the generator will be

$$L_n^* h = \frac{\partial}{\partial x}(h\frac{x}{2\eta^2}) + \frac{\partial}{\partial \eta}(h(q\eta - \frac{g_n(x)}{\sqrt{2\pi}})) + \frac{\partial^2}{\partial x^2}(h\frac{1}{\eta^2}),$$

for $h(\cdot, \cdot)$ belonging to the domain of L_n^*.

Let $\pi_n(\cdot, \cdot)$ denote the density corresponding to the invariant distribution of $\{(X_{n,t}, \eta_{n,t})\}$. Then it satisfies the equation

$$\int (L_n^* \pi_n(x, \eta)) f(x, \eta) dx d\eta \;=\; 0, \tag{4.2.35}$$

for $f(\cdot, \cdot)$ belonging to the domain of L_n which contains the class of infinitely differentiable and compactly supported functions. Therefore the above equation is equivalent to

$$L_n^* \pi_n(x, \eta) \;=\; 0.$$

Writing $\pi_n^{(i)}(\cdot,\cdot)$ as the first derivative w.r.t the i^{th} component, $i \in 1,2$ and $\pi_n^{(i,j)}(\cdot,\cdot)$ as the second order derivatives we have the above equation as

$$\frac{1}{2\eta^2}\pi_n(x,\eta) + \frac{x}{2\eta^2}\pi_n^{(1)}(x,\eta) + (q\eta - \frac{g_n(x)}{\sqrt{2\pi}})\pi_n^{(2)}(x,\eta) + q\pi_n(x,\eta) + \frac{1}{2\eta^2}\pi_n^{(2,0)}(x,\eta) \quad = \quad 0.$$

Multiplying by $2\eta^2$ on both sides of the above and taking integrals w.r.t x from 0 to M and η from 0 to N (which we do not write explicitly) we have

$$\begin{aligned}\int\int \pi_n(x,\eta)dxd\eta + \int\int x\pi_n^{(1)}(x,\eta)dxd\eta + \int\int 2\eta^2(q\eta - \frac{g_n(x)}{\sqrt{2\pi}})\pi_n^{(2)}(x,\eta)dxd\eta \quad &+ \\ q\int\int 2\eta^2\pi_n(x,\eta)dxd\eta + \int\int \pi_n^{(2,0)}(x,\eta)dxd\eta \quad &= \quad 0.\end{aligned} \tag{4.2.36}$$

For the second term we apply integration by parts to get

$$\int\int x\pi_n^{(1)}(x,\eta)dxd\eta \quad = \quad \int_0^N M\pi_n(M,\eta)d\eta - \int_0^N\int_0^M \pi_n(x,\eta)dxd\eta.$$

And similarly the other terms are

$$\begin{aligned}\int\int 2\eta^3 q\pi_n^{(2)}(x,\eta)dxd\eta \quad &= \quad 2q\Big(\int_0^M N^3\pi_n(x,N)dx - 3\int_0^M\int_0^N \eta^2\pi_n(x,\eta)d\eta dx\Big) \\ -\int\int 2\eta^2\frac{g_n(x)}{\sqrt{2\pi}}\pi_n^{(2)}(x,\eta)dxd\eta \quad &= \quad -\sqrt{\frac{2}{\pi}}\Big(\int_0^M N^2\pi_n(x,N)g_n(x)dx \\ &- \quad 2\int_0^M g_n(x)\int_0^N \eta\pi_n(x,\eta)d\eta dx\Big)\end{aligned}$$

The fourth term $\int\int 2\eta^2\pi_n(x,\eta)dxd\eta$ is bounded. Assuming the interchange of integral and differential sign and taking $N \to \infty$ we have the last term as

$$\int_0^\infty\int_0^\infty \pi_n^{(2)}(x,\eta)dxd\eta = \int_0^\infty \pi_n^{(1)}(M,\eta)d\eta = \pi_n^{(1)}(M)$$

since $\pi_n(x,\eta)$ is symmetric w.r.t. x around 0 for each fixed η and hence $\pi_n^{(1)}(x,\eta)$ is an odd function and hence $\pi_n^{(1)}(0,\eta) = 0$ for each fixed η. We further assume that $\pi_n^{(1)}(x,0) = 0$

for each x. Thus, as $N \to \infty$ we have

$$\begin{aligned}\pi_n^{(1)}(M) &+ \int_0^\infty M\pi_n(M,\eta)d\eta + 2q\Big(\lim_{N\to\infty}\int_0^M N^3\pi_n(x,N)dx - 3\int_0^M\int_0^\infty \eta^2\pi_n(x,\eta)dxd\eta\Big)\\ &- \sqrt{\frac{2}{\pi}}\Big(\lim_{N\to\infty}\int_0^M N^2\pi_n(x,N)g_n(x)dx - 2\int_0^M g_n(x)\int_0^\infty \eta\pi_n(x,\eta)dxd\eta\Big)\\ &+ 2q\int_0^M\int_0^\infty \eta^2\pi_n(x,\eta)d\eta dx = 0 \qquad (4.2.37)\end{aligned}$$

In Remark 20 we have argued that the m.g.f of $\pi_n(x,\eta)$ is finite for all $s_1, s_2 \in \mathbb{R}$ are uniformly bounded in n. This proves that the fourth, sixth and seventh term on the LHS of Equation (4.2.37) are bounded uniformly in n. For the same reason the other terms can also be assumed to be bounded uniformly in n except under those pathological functions which are integrable but the $\limsup$ is not finite. In view of of Remark 20 this may be a reasonable assumption. We argue in a similar fashion for negative M. This is sufficient for the X-marginal of $\pi_n(\cdot,\cdot)$ to be equicontinuous over compact subsets of $\mathbb{R}$. In a similar manner uniform boundedness of $\pi_n(\cdot)$ (over n) may also be assumed in view of Remark 20.

Remark 22. It should be noted that in Section 4.2.3 we prove the convergence of the marginal of the limiting distribution of(X_t, η_t) by the method of moments. Therefore it is not necessary to assume that the density of (X_t, η_t) for each t exists and consequently the joint density of the invariant distribution.

4.2.3 Identifying the limiting distribution

We first prove a lemma that will be required in this subsection. For any $s > 0$, define $F_s(t) = s\int\limits_0^t \theta_u^2 du$.

Lemma 19.

$$\lim_{t\to\infty} E(e^{-F_s(t)}) = 0, \ \forall s > 0.$$

Proof: We prove for $s = 1$. The proof can be carried out in a similar fashion for any $s > 0$.

$$\frac{F_1(t)}{t} = \frac{1}{t}\int_0^t \theta_s^2 ds \geq \frac{1}{\frac{1}{t}\int_0^t \frac{1}{\theta_s^2} ds} = \frac{1}{\frac{1}{t}\int_0^t \eta_s^2 ds}, \tag{4.2.38}$$

where the last but one inequality follows from Jensen's (by taking $\psi(x) = \frac{1}{x},\ x > 0$ which is convex). This implies

$$\frac{1}{\frac{1}{t}F_1(t)} \leq \frac{1}{t}\int_0^t \eta_s^2 ds \Rightarrow \frac{1}{F_1(t)} \leq \frac{1}{t}\frac{1}{t}\int_0^t \eta_s^2 ds. \tag{4.2.39}$$

Therefore,

$$\begin{aligned}
e^{-F_1(t)} &= \frac{1}{e^{F_1(t)}} \leq \frac{1}{F_1(t)} \text{ (since } e^x \geq x,\ \forall x > 0) \\
&\leq \frac{1}{t}\frac{1}{t}\int_0^t \eta_s^2 ds \text{ (from (4.2.39))} \\
\Rightarrow E(e^{-F_1(t)}) &\leq \frac{1}{t}E\Big(\frac{1}{t}\int_0^t \eta_s^2 ds\Big) \leq \frac{1}{t}C.
\end{aligned}$$

where $C = \sup_{t>0} E(\frac{1}{t}\int_0^t \eta_s^2 ds) < \infty$, from Lemma 15. So

$$\lim_{t\to\infty} E(e^{-F_1(t)}) = 0.$$

■

LEMMA 20. Assuming that all the moments of X_0 and θ_0 exists we have

$$\lim_{t\to\infty} E(X_t^r) = \begin{cases} \frac{(2k)!}{2^k k!} & \text{when } r = 2k \\ 0 & \text{when } r = 2k+1. \end{cases}$$

Proof: We prove using induction for both even and odd moments:

Even moments

1. We first show $\lim_{t\to\infty} E(X_t^2) = 1$.
   Applying Itô's lemma to X_t^2 we have

$$dX_t^2 = \Big(-X_t^2\theta_t^2 + \theta_t^2\Big)dt + 2X_t\theta_t dW_t.$$

Multiplying by the integrating factor $e^{F_1(t)}$, where $F_1(t) = \int_0^t \theta_s^2 ds$, on both sides of the above equation we have

$$\begin{aligned}
d\Big(X_t^2 e^{F_1(t)}\Big) &= e^{F_1(t)}\theta_t^2 dt + e^{F_1(t)} X_t\theta_t dW_t \\
\Rightarrow X_t^2 &= e^{-F_1(t)}[X_0^2 + \int_0^t e^{F_1(s)}\theta_s^2 ds + 2\int_0^t e^{F_1(s)} X_s\theta_s dW_s] \\
&= e^{-F_1(t)}[X_0^2 + \int_0^t d(e^{F_1(s)}) + 2\int_0^t e^{F_1(s)} X_s\theta_s dW_s] \\
&= e^{-F_1(t)}[X_0^2 + e^{F_1(t)} - 1 + 2\int_0^t e^{F_1(s)} X_s\theta_s dW_s] \\
&= X_0^2 e^{-F_1(t)} + 1 - e^{-F_1(t)} + 2\int_0^t e^{F_1(s)-F_1(t)} X_s\theta_s dW_s \\
\Rightarrow E(X_t^2) &= E(e^{-F_1(t)})E(X_0^2) + 1 - E(e^{-F_1(t)}) \\
&+ 2E\Big(e^{-F_1(t)}\int_0^t e^{F_1(s)} X_s\theta_s dW_s\Big).
\end{aligned}$$

From the proof of Lemma 12 we have that the third expectation is zero (by substituting $m = 1$). Therefore

$$\begin{aligned}
E(X_t^2) &= E(e^{-F_1(t)})E(X_0^2) + 1 - E(e^{-F_1(t)}) \\
\Rightarrow \lim_{t\to\infty} E(X_t^2) &= E(X_0^2)\lim_{t\to\infty} E(e^{-F_1(t)}) && (4.2.40) \\
&+ 1 - \lim_{t\to\infty} E(e^{-F_1(t)}) && (4.2.41)
\end{aligned}$$

Now $\lim_{t\to\infty} E(e^{-F_k(t)}) = 0$ by Lemma 19. Therefore,

$$\lim_{t\to\infty} E(X_t^2) = 1 \text{ from } (4.2.41).$$

2. Assume this holds for $1 \le m \le k-1$, i.e.,

$$\lim_{t\to\infty} E(X_t^{2m}) = \frac{(2m)!}{2^m m!} \quad \text{for } 1 \le m \le k-1.$$

3. From Itô's lemma applied to X_t^{2k}

$$dX_t^{2k} \;=\; \Big(-kX_t^{2k}\theta_t^2 + k(2k-1)X_t^{2k-2}\theta_t^2\Big)dt + 2kX_t^{2k-1}\theta_t dW_t.$$

Multiplying with the integrating factor $e^{F_k(t)}$ on both sides of the above equation and rearranging we have that

$$\begin{aligned}
d\Big(X_t^{2k}e^{F_k(t)}\Big) &= k(2k-1)e^{F_k(t)}X_t^{2k-2}\theta_t^2 dt + 2ke^{F_k(t)}X_t^{2k-1}\theta_t dW_t \\
\Rightarrow X_t^{2k} &= e^{-F_k(t)}[X_0^{2k} + (2k-1)\int_0^t ke^{F_k(s)}X_s^{2k-2}\theta_s^2 ds \\
&+ 2k\int_0^t e^{F_k(s)}X_s^{2k-1}\theta_s dW_s \\
\Rightarrow E(X_t^{2k}) &= E(e^{-F_k(t)})E(X_0^{2k}) \\
&+ (2k-1)E(\int_0^t ke^{-F_k(t)}e^{F_k(s)}X_s^{2k-2}\theta_s^2 ds) \\
&+ E\Big(2e^{-F_k(t)}\int_0^t ke^{F_k(s)}X_s^{2k-1}\theta_s dW_s\Big).
\end{aligned} \tag{4.2.42}$$

We have proved in Lemma 12 that the third expectation in the RHS of (4.2.42) is zero (by substituting m = k). Writing

$$\begin{aligned}
A_{k,2m-2}(t) &:= E(e^{-F_k(t)}k\int_0^t e^{F_k(s)}X_s^{2m-2}\theta_s^2 ds) \\
&= E(e^{-F_k(t)}\int_0^t X_s^{2m-2}d(e^{F_k(s)})), \quad \text{for } 1 \le m \le k
\end{aligned} \tag{4.2.43}$$

we have,

$$
\begin{aligned}
E(X_t^{2k}) &= E(X_0^{2k})E(e^{-F_k(t)}) \\
&+ (2k-1)A_{k,2k-2}(t).
\end{aligned}
\tag{4.2.44}
$$

Now by the integration by parts we have,

$$
\begin{aligned}
\int_0^t X_s^{2m} d(e^{F_k(s)}) &= X_t^{2m} e^{F_k(t)} - X_0^{2m} - \int_0^t e^{F_k(s)} d(X_s^{2m}) \\
&= X_t^{2m} e^{F_k(t)} - X_0^{2m} - \int_0^t e^{F_k(s)} \Big((-mX_s^{2m}\theta_s^2 \\
&+ m(2m-1)X_s^{2m-2}\theta_s^2)ds + \int_0^t 2mX_s^{2m-1}\theta_s dW_s \Big),
\end{aligned}
$$

using

$$
dX_t^{2m} = -mX_t^{2m}\theta_t^2 dt + m(2m-1)X_t^{2m-2}\theta_t^2 dt + 2mX_t^{2m-1}\theta_t dW_t.
$$

Therefore multiplying by $e^{-F_k(t)}$ on both sides of the above equation we have

$$
\begin{aligned}
e^{-F_k(t)} \int_0^t X_s^{2m} d(e^{F_k(s)}) &= e^{-F_k(t)} \int_0^t k\theta_s^2 e^{F_k(s)} X_s^{2m} ds \\
&= X_t^{2m} - X_0^{2m} e^{-F_k(t)} \\
&+ e^{-F_k(t)} \int_0^t m e^{F_k(s)} X_s^{2m} \theta_s^2 ds \\
&- e^{-F_k(t)} \int_0^t m(2m-1) e^{F_k(s)} X_s^{2m-2} \theta_s^2 ds \\
&+ 2m e^{-F_k(t)} \int_0^t X_s^{2m-1} e^{F_k(s)} \theta_s dW_s.
\end{aligned}
$$

Taking expectations on both sides and recalling the definition of $A_{k,2m}(t)$ from (4.2.43) we have ,

$$\begin{aligned} A_{k,2m}(t) &= E(X_t^{2m}) - E(e^{-F_k(t)})E(X_0^{2m}) + \frac{m}{k}A_{k,2m}(t) \\ &- \frac{m(2m-1)}{k}A_{k,2m-2}(t) + 0. \end{aligned} \tag{4.2.45}$$

That the last expectation is zero follows from Lemma 12. This implies that

$$\begin{aligned} (1-\frac{m}{k})A_{k,2m}(t) &= E(X_t^{2m}) - E(e^{-F_k(t)})E(X_0^{2m}) \\ &- \frac{m(2m-1)}{k}A_{k,2m-2}(t). \end{aligned} \tag{4.2.46}$$

Now,

$$\begin{aligned} A_{k,0}(t) &= E(e^{-F_k(t)}\int_0^t k\theta_s^2 e^{F_k(s)}ds) \\ &= E(e^{-F_k(t)}\int_0^t d(e^{F_k(s)})) = 1 - e^{-F_k(t)}. \end{aligned} \tag{4.2.47}$$

Define $B_{k,2m} = \lim_{t\to\infty} A_{k,2m}(t)$ (when the limit exists). Taking limits as $t \to \infty$ on both sides of (4.2.47) and applying Lemma 19 we get:

$$B_{k,0} = 1 - \lim_{t\to\infty} e^{-F_k(t)} = 1. \tag{4.2.48}$$

Hence $\lim\limits_{t\to\infty} A_{k,2m}(t)$ exists for $m = 0$.
Taking $m = 1, 2, 3, \ldots, k-1$ in (4.2.46) we get that $\lim\limits_{t\to\infty} A_{k,2m}(t)$ exists, since

$$\begin{aligned} (1-\frac{m}{k})\lim_{t\to\infty} A_{k,2m}(t) &= \lim_{t\to\infty} E(X_t^{2m}) - \frac{m(2m-1)}{k}\lim_{t\to\infty} A_{k,2m-2}(t) \\ \Rightarrow B_{k,2m} &= \frac{k}{k-m}\lim_{t\to\infty} E(X_t^{2m}) - \frac{m(2m-1)}{k-m}B_{k,2m-2}. \end{aligned} \tag{4.2.49}$$

Substituting different values of $m = 0, 1, 2, \ldots, k-1$ in (4.2.49) and applying induction hypothesis, that $\lim\limits_{t\to\infty} E(X_t^{2m}) = \frac{(2m)!}{2^m m!}$, for $0 \le m \le k-1$, we get:

$$
\begin{aligned}
B_{k,0} &= 1 \\
B_{k,2} &= \frac{k}{k-1}1 - \frac{1}{k-1}1 = 1 \\
B_{k,4} &= \frac{k}{k-2}3 - \frac{2.3}{k-2}1 = 3 \\
B_{k,6} &= \frac{k}{k-3}5.3 - \frac{3.5}{k-3}3 = 5.3 \\
B_{k,8} &= \frac{k}{k-4}7.5.3 - \frac{4.7}{k-4}5.3 = 7.5.3 \\
\ldots & \\
B_{k,2k-2} &= k(2k-3)(2k-5)\ldots 3.1 - (k-1)(2k-3)B_{k,2k-4} \\
&= k(2k-3)(2k-5)\ldots 3.1 - (k-1)(2k-3)\ (2k-5)\ldots 3.1 \\
&= (2k-3)(2k-5)\ldots 3.1(k-k+1) \\
&= \frac{(2k-2)!}{2^{k-1}(k-1)!}.
\end{aligned}
$$

Therefore applying Lemma 19 to Equation (4.2.44) :

$$
\begin{aligned}
\lim_{t\to\infty} E(X_t^{2k}) &= (2k-1)B_{k,2k-2} \\
&= (2k-1)\frac{(2k-2)!}{2^{k-1}(k-1)!} = \frac{2k(2k-1)!}{2^k k!} \\
&= \frac{(2k)!}{2^k k!}.
\end{aligned}
\tag{4.2.50}
$$

Odd moments

1. To find the odd moments of X_t we perform similar procedure as above. We have

$$
dX_t = -X_t \frac{\theta_t^2}{2} dt + \theta_t dW_t \tag{4.2.51}
$$

Define $G_k(t) = \frac{2k+1}{2}\int\limits_0^t \theta_s^2 ds, k \in \mathbb{N} \cup \{0\}$. Multiply by the integrating factor $e^{G_0(t)}$ on both sides of (4.2.51) and rearrange to get

$$\begin{aligned} d(e^{G_0(t)}X_t) &= e^{G_0(t)}\theta_t dW_t \\ \Rightarrow X_t &= X_0 e^{-G_0(t)} + e^{-G_0(t)}\int\limits_0^t e^{G_0(s)}\theta_s dW_s \\ \Rightarrow E(X_t) &= E(X_0)E(e^{-G_0(t)}) + E\Big(e^{-G_0(t)}\int\limits_0^t e^{G_0(s)}\theta_s dW_s\Big). \end{aligned} \tag{4.2.52}$$

From Lemma 19 we have

$$\lim_{t\to\infty} E(e^{-G_0(t)}) = 0.$$

Therefore from (4.2.52) we have

$$\lim_{t\to\infty} E(X_t) = 0.$$

2. Let $k \geq 1$ be any positive integer. Assume that

$$\lim_{t\to\infty} E(X_t^{2m-1}) = 0 \text{ where } \ m = 1, 2, \ldots, k.$$

3. Applying Itô's lemma to X_t^{2k+1} we get

$$\begin{aligned} dX_t^{2k+1} &= (2k+1)X_t^{2k}dX_t + \frac{1}{2}(2k+1)2kX_t^{2k-1}\theta_t^2 dt \\ &= (2k+1)X_t^{2k}\Big(-X_t\frac{\theta_t^2}{2}dt + \theta_t dW_t\Big) + (2k+1)k\theta_t^2 X_t^{2k-1}dt \\ &= \Big(-\frac{1}{2}(2k+1)X_t^{2k+1}\theta_t^2 \\ &+ (2k+1)kX_t^{2k-1}\theta_t\Big)dt + (2k+1)\theta_t X_t^{2k} dW_t. \end{aligned} \tag{4.2.53}$$

Multiplying by the integrating factor $e^{G_k(t)}$ on both sides of (4.2.53) and rearranging we get:

$$\begin{aligned} d\Big(X_t^{2k+1}e^{G_k(t)}\Big) &= k(2k+1)e^{G_k(t)}\theta_t^2 X_t^{2k-1}dt + (2k+1)e^{G_k(t)}\theta_t X_t^{2k}dW_t \\ \Rightarrow X_t^{2k+1} &= e^{-G_k(t)}\Big[X_0^{2k+1} + k(2k+1)\int_0^t e^{G_k(s)}\theta_s^2 X_s^{2k-1}ds \\ &+ (2k+1)\int_0^t e^{G_k(s)}\theta_s X_s^{2k}dW_s\Big]. \end{aligned}$$

Thus

$$\begin{aligned} E(X_t^{2k+1}) &= E(e^{-G_k(t)})E(X_0^{2k+1}) \\ &+ E\Big(k(2k+1)e^{-G_k(t)}\int_0^t e^{G_k(s)}\theta_s^2 X_s^{2k-1}ds\Big) \\ &+ (2k+1)E\Big(e^{-G_k(t)}\int_0^t e^{G_k(s)}X_s^{2k}\theta_s dW_s\Big). \end{aligned} \tag{4.2.54}$$

From Lemma 13 we have the third expectation is zero. That is

$$E\Big(e^{-G_k(t)}\int_0^t e^{G_k(s)}X_s^{2k}\theta_s dW_s\Big) = 0.$$

Defining

$$\begin{aligned} C_{k,2m-1}(t) &:= E\Big(k(2k+1)e^{-G_k(t)}\int_0^t e^{G_k(s)}\theta_s^2 X_s^{2m-1}ds\Big) \\ &= E\Big(2ke^{-G_k(t)}\int_0^t X_s^{2m-1}d(e^{G_k(s)})\Big). \end{aligned} \tag{4.2.55}$$

We have from (4.2.54).

$$E(X_t^{2k+1}) = E(e^{-G_k(t)})E(X_0^{2k+1}) + C_{k,2k-1}(t). \tag{4.2.56}$$

Now by integration by parts

$$\int_0^t X_s^{2m-1} d(e^{G_k(s)}) = X_t^{2m-1} e^{G_k(t)} - X_0^{2m-1} - \int_0^t e^{G_k(s)} d(X_s^{2m-1}).$$

Applying Itô's lemma to X_t^{2m-1} we have

$$\begin{aligned} dX_t^{2m-1} &= (2m-1)X_t^{2m-2}dX_t + (2m-1)(m-1)X_t^{2m-3}\theta_t^2 dt \\ &= -\frac{2m-1}{2}X_t^{2m-1}\theta_t^2 dt + (m-1)(2m-1)X_t^{2m-3}\theta_t^2 dt + (2m-1)X_t^{2m-2}\theta_t dW_t. \end{aligned}$$

Substituting in the above equation we have

$$\begin{aligned} \int_0^t X_s^{2m-1} d(e^{G_k(s)}) &= X_t^{2m-1} e^{G_k(t)} - X_0^{2m-1} + \int_0^t (2m-1) e^{G_k(s)} X_s^{2m-1} \frac{\theta_s^2}{2} ds \\ &- (2m-1)(m-1)\int_0^t e^{G_k(s)} X_s^{2m-3}\theta_s^2 ds - \int_0^t (2m-1) e^{G_k(s)} X_s^{2m-2}\theta_s dW_s. \end{aligned} \tag{4.2.57}$$

Multiplying both sides by $e^{-G_k(t)}$, taking expectations in (4.2.57) and recalling the definition of $C_{k,2m-1}$ from (4.2.55) we have

$$\begin{aligned} C_{k,2m-1}(t) &= E\Big(2ke^{-G_k(t)}\int_0^t X_s^{2m-1} d(e^{G_k(s)})\Big) \\ &= 2kE(X_t^{2m-1}) - 2kE(e^{-G_k(t)})E(X_0^{2m-1}) + \frac{2m-1}{(2k+1)}C_{k,2m-1}(t) \\ &- \frac{(2m-1)(2m-2)}{(2k+1)}C_{k,2m-3}(t) - (2m-1)E\Big(e^{-G_k(t)}\int_0^t e^{G_k(s)} X_s^{2m-2}\theta_s dW_s\Big). \end{aligned} \tag{4.2.58}$$

Now by Lemma 13 where it is shown that

$$E\Big(e^{-G_k(t)}\int_0^t \theta_s e^{G_k(s)} X_s^{2m} dW_s\Big) = 0, \text{ for } 0 \le m \le k-1,$$

we have that the third expectation is zero. Now

$$
\begin{aligned}
C_{k,1}(t) &= 2kE\Big(e^{-G_k(t)}\int_0^t X_s d(e^{G_k(s)})\Big)\\
&= 2kE\Big(e^{-G_k(t)}\Big(X_t e^{G_k(t)} - X_0 e^{G_k(0)} - \int_0^t e^{G_k(s)} dX_s\Big)\Big)\\
&= 2kE\Big(X_t - X_0 e^{-G_k(t)} - e^{-G_k(t)}\int_0^t e^{G_k(s)} dX_s\Big) \text{ since } G_k(0) = 0.
\end{aligned}
$$

From the SDE of X_t we have

$$
\begin{aligned}
C_{k,1}(t) &= 2kE\Big(X_t - X_0 e^{-G_k(t)} - e^{-G_k(t)}\int_0^t e^{G_k(s)}\Big(-X_s\frac{\theta_s^2}{2}ds + \theta_s dW_s\Big)\Big)\\
&= 2k\Big(E(X_t) - E(X_0 e^{-G_k(t)}) + \frac{1}{2}E\Big(e^{-G_k(t)}\int_0^t e^{G_k(s)}\theta_s^2 X_s ds\Big)\\
&- E\Big(e^{-G_k(t)}\int_0^t e^{G_k(s)}\theta_s dW_s\Big)\Big)\\
&= 2k\Big(E(X_t) - E(X_0 e^{-G_k(t)})\Big) + \frac{1}{2k+1}C_{k,1}(t) + 0,
\end{aligned}
$$

since $E\Big(e^{-G_k(t)}\int_0^t e^{G_k(s)}\theta_s dW_s\Big) = 0$ from Lemma 13 (by substituting m = 0). Therefore

$$
(1 - \frac{1}{2k+1})C_{k,1}(t) = 2k\Big(E(X_t) - E(X_0 e^{-G_k(t)})\Big). \tag{4.2.59}
$$

Now we have proved that $\lim_{t\to\infty} E(X_t) = 0 = \lim_{t\to\infty} E(e^{-G_k(t)})$. Defining $D_{k,m} = \lim_{t\to\infty} C_{k,m}(t)$ for $m = 1, 3, 5, \ldots, 2k-1$, wherever it exists, we have from (4.2.59)

$$
D_{k,1} = 0.
$$

From (4.2.58) we have

$$\begin{aligned}(1-\frac{2m-1}{2k+1})C_{k,2m-1}(t) &= 2kE(X_t^{2m-1})-2kE(e^{-G_k(t)})E(X_0^{2m-1}) \\ &- \frac{(2m-1)(2m-2)}{(2k+1)}C_{k,2m-3}(t). \qquad (4.2.60)\end{aligned}$$

By induction hypothesis $\lim_{t\to\infty} E(X_t^{2m-1}) = 0$ for $m = 1, 2, \dots, k$. Since $D_{k,1} = 0$ from (4.2.60) we have by iteration $\lim_{t\to\infty} C_{k,2m-1}(t)$ exists and equals to 0 for $m = 1, 2, \dots, k$, i.e.

$$D_{k,j} = 0 \ \text{ for } j = 1, 3, \dots, 2k-1.$$

Therefore, from 4.2.54 we have that

$$\lim_{t\to\infty} E(X_t^{2k+1}) = 0. \qquad (4.2.61)$$

Thus combining (4.2.50) and (4.2.56) we see that the limiting moments of $\{X_s\}$ matches with that of a $N(0,1)$ distribution. Since the limiting distribution admits a smooth density, invoking uniqueness of moment generating function we can infer that the limiting distribution of $\{X_s\}$ is $N(0,1)$. This completes the proof of Theorem 4. ■

Remark 23. From (4.2.17) we have θ_t satisfying the equation

$$\begin{aligned}\theta_t &= \frac{e^{qt}}{\eta_0+\frac{1}{\sqrt{2\pi}}\int_0^t e^{qs}|X_s|ds} \\ \Rightarrow \theta_t^2 &= \frac{e^{2qt}}{\left(\eta_0+\frac{1}{\sqrt{2\pi}}\int_0^t e^{qs}|X_s|ds\right)^2} \\ &\geq \frac{e^{2qt}}{2\left(\eta_0^2+\frac{1}{2\pi}(\int_0^t e^{qs}|X_s|ds)^2\right)} \\ &= \frac{e^{2qt}}{2\eta_0^2+\frac{(e^{qt}-1)^2}{\pi q^2}(\int_0^t \frac{q}{e^{qt}-1}e^{qs}|X_s|ds)^2} \\ &\geq \frac{e^{2qt}}{2\eta_0^2+\frac{e^{qt}-1}{\pi q}\int_0^t e^{qs}|X_s|^2 ds}, \qquad (4.2.62)\end{aligned}$$

where the last inequality follows from the fact that

$$\Big(\int_0^t \frac{q}{e^{qt}-1} e^{qs}|X_s| ds\Big)^2 \;\le\; \int_0^t \frac{q}{e^{qt}-1} e^{qs}|X_s|^2 ds.$$

This is true by the Jensen's inequality

$$(E|X_s|)^2 \;\le\; E(|X_s|^2),$$

with the expectation computed with respect to the density $f(x) = \frac{q}{e^{qt}-1}e^{qx}, 0 < x < t$, for any $t > 0$. Therefore

$$\begin{aligned} E(\theta_t^2) &\ge \frac{e^{2qt}}{2E(\eta_0^2) + \frac{e^{qt}-1}{\pi q}\int_0^t e^{qs}E(X_s^2)ds} \\ &\ge \frac{e^{2qt}}{2E(\eta_0^2) + \frac{(e^{qt}-1)^2(1+E(X_0^2))}{\pi q^2}}, \end{aligned}$$

where the last inequality follows from (4.2.40) that

$$E(X_t^2) \le 1 + E(X_0^2) \;\forall t > 0.$$

Therefore

$$\begin{aligned} \liminf_{t\to\infty} E(\theta_t^2) &\ge \liminf_{t\to\infty} \frac{e^{2qt}}{2E(\eta_0^2) + \frac{(e^{qt}-1)^2(1+E(X_0^2))}{\pi q^2}} = \frac{\pi q^2}{1+E(X_0^2)} \\ \Rightarrow \liminf_{t\to\infty} \frac{1}{t}\int_0^t E(\theta_u^2)du &\ge \frac{\pi q^2}{1+E(X_0^2)}, \quad \text{by Fatou's lemma.} \end{aligned}$$

In particular if $X_0 = 0$ almost surely, then

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t E(\theta_u^2)du \;\ge\; \pi q^2.$$

This gives a lower bound to the growth of θ_t.

REMARK 24. **Rates of convergence of Adaptive and Standard MCMC:** Recalling the SDE for AMCMC for Normal target density for X_t is given as:

$$dX_t = -X_t \frac{\theta_t^2}{2} dt + \theta_t dW_t.$$

Multiplying by the integrating factor and performing the usual operations we get:

$$E(X_t) = E(X_0 e^{-G_t}), \text{ where } G_t = \int_0^t \frac{\theta_s^2}{2} ds. \quad (4.2.63)$$

Similar equation for the SMCMC Y_t is:

$$dY_t = -Y_t \frac{\theta_0^2}{2} dt + \theta_0 dW_t.$$

Applying similar computations we get

$$E(Y_t) = E(Y_0 e^{-\tilde{G}_t}), \text{ where } \tilde{G}_t = \int_0^t \frac{\theta_0^2}{2} ds = \frac{\theta_0^2}{2} t.$$

Similar computation with X_t^2 will give (see the proof of Lemma 20)

$$E(X_t^2) = E(X_0^2 e^{-G(t)}) + 1 - E(e^{-G(t)}). \quad (4.2.64)$$

It is therefore clear from the Equations (4.2.63) and (4.2.64) that the quantity regulating the speed to convergence is $G(t)$ (or $\tilde{G}(t)$). The faster $G(t)$ (or $\tilde{G}(t)$) goes to ∞, the faster the chain converges to its invariant distribution (which is standard Normal in this case). For SMCMC the rate is exponential in t. For AMCMC it depends on the behaviour of $\int_0^t \theta_s^2 ds$. We have shown in Lemma 17 that $\limsup_{t\to\infty} \frac{1}{t} \int_0^t E(\theta_u^2) du < C < \infty$ for any $k \in \mathbb{N}$ when the target distribution is standard Normal. Combining this with Remark 23 we find that the rate of convergence of the AMCMC will be exponentially going to zero at a rate which is linear in t. Thus the comparison between the rate of convergence of AMCMC and SMCMC will depend on the lower bound πq^2 and the upper bound C (as in Lemma 17 for $k = 4$) and θ_0^2. If the bound can be obtained in the almost sure sense, and not in the L_1 sense then it might be possible to directly compare SMCMC and AMCMC.

REMARK 25. It is true that for the discrete time SMCMC, higher value of θ_0 will delay convergence to stationarity of the chain. However from the diffusion approximation to the discrete SMCMC we see that the rate of convergence is determined by θ_0 - a higher value of θ_0 will imply a faster convergence to stationarity. For the AMCMC situations we have a different situations. The simulations in Figures 3.1 and 3.2 show that the trajectories of θ_t converge for large values of t. This is in tune to our theoretical findings that for a standard Normal target with standard Normal proposals, the time average moments of θ_t are bounded. Since this happens for any starting value of θ_0, we recommend that this limiting value should be used for selecting the optimal value of θ_0. One should run the AMCMC sufficiently long, till the point where θ_t changes no further. From that point onwards one should keep the level of θ_t same and run a simple SMCMC.

4.3 Summary

Verifying the conditions of Roberts *et al.* (see Proposition 3, Chapter 3) for checking the ergodicity of an AMCMC can sometimes prove to be difficult. In Chapter 3, we considered an AMCMC with the proposal kernel dependent on the previously generated sample and an arbitrary target distribution. There we performed a diffusion approximation technique to look at the continuous time version of the discrete chain. In this chapter we narrowed down to the case where the target distribution is standard Normal. We investigate whether the invariant distribution of the diffusion is indeed the target distribution. It turns out that the resulting diffusion (which although singular) admits a unique invariant distribution. Then computing the limiting moments (both even and odd) of X_t we identify the limiting distribution to be $N(0, 1)$.
The techniques applied here are specific only when the target distribution is Normal. We hope that this can also be extended to other target distributions, where an identification of the limiting moments is possible. Also more choices of the proposal distribution can be made, where the kernel is dependent on a finite (or possibly infinite) past. Some issues in the choice of the proposal and target are dealt in Chapter 5.

Chapter 5

Diffusion Approximation for general target and proposal distribution

5.1 Introduction

In the earlier chapter we were concerned with the situation when the target and proposal distribution are both standard univariate Normal. In many situations the standard Normal is the choice as a proposal, since generating samples from it is easy (for example, using the Box Muller technique). Also in the proof of the diffusion approximation it requires the existence of the first two moments of the proposal. A natural question is how can the result in Chapter 4 be extended for general target and proposal distributions. We try to address these issues in this chapter.
In the diffusion approximation with the Normal proposal (see Remark 12 of Theorem 3 of Chapter 3) we have the non-explosive condition involving the derivative of density of the target distribution $\psi(\cdot)$ see, Equation (5.2.1). In Section 5.2.1 we consider sub-classes of densities satisfying the condition, see Equation (5.2.2) to 5.2.4. We show that this corresponds to the existence or non existence of the m.g.f of the target $\psi(\cdot)$ in the whole of $\mathbb{R}$ or in a neighborhood of zero. In Section 5.2.2 we show that if the target $\psi(\cdot)$ satisfy condition 5.2.2 and (5.2.3) with $\alpha \in [1, 2)$ then the limiting distribution of the diffusion corresponding to the AMCMC also share the same property, see Theorem 6 and 7 for the

precise statement. In Section 5.2.4 we look at the standard Cauchy as the target distribution in Theorem 9. We show that the diffusion approximation method cannot be used for simulation when the standard Cauchy is the target. The multivariate scenario is dealt in Section 5.3 where we obtain the diffusion of the multivariate AMCMC. In Section 5.6 we obtain the diffusion approximation for light tailed proposals. We also explain why heavy tail choices (such as the Cauchy distribution) of the proposal distribution will not work. Specifically, we investigate what goes wrong when we look at the ϵ localized infinitesimal drift and diffusion coefficient when the proposal distribution is Cauchy. We heuristically relate the observation to a Theorem in Chapter 2.

REMARK 26. In this chapter we are concerned only with stability in the sense of distributions, i.e., whether the target distribution is same as the limiting distribution of the SDE. For other concepts of stability of Markov process (topological and probabilistic) see Meyn and Tweedie, [46].

5.2 General Target distribution

5.2.1 Types of target densities

First we recall the SDE of corresponding to a general target distribution $\psi(\cdot)$ given by Equation (3.4.1) in Chapter 3.

$$\begin{aligned} dX_t &= \frac{\theta_t^2}{2}\frac{\psi'(X_t)}{\psi(X_t)}dt + \theta_t dW_t \\ d\theta_t &= \theta_t\Big(q - \frac{1}{\sqrt{2\pi}}\theta_t\frac{|\psi'(X_t)|}{\psi(X_t)}\Big)dt. \end{aligned}$$

We consider target distribution with linear growth condition as in Remark 12, i.e.,

$$\frac{|\psi'(x)|}{\psi(x)} \leq a|x| + b, \;\; a \geq 0, b \geq 0, x \in \mathbb{R}. \tag{5.2.1}$$

Such condition ensures that the solution of the SDE is non-explosive, see, Remark 13 of Chapter 3 target density). Under this hypothesis we work with the following three classes of target distributions. They are

$$x\frac{\psi'(x)}{\psi(x)} \leq -a|x|^2 + b,\ a > 0,\ b \geq 0, \tag{5.2.2}$$

$$x\frac{\psi'(x)}{\psi(x)} \leq -a_2|x|^\alpha + b_2,\ a_2 > 0, b_2 \geq 0, \alpha \in (0,2),\ \text{and}, \tag{5.2.3}$$

$$x\frac{\psi'(x)}{\psi(x)} \leq -a_3,\ a_3 > 0, \tag{5.2.4}$$

THEOREM 5. If the density $\psi(\cdot)$ satisfies the condition in Equation (5.2.2) then the m.g.f of $\psi(\cdot)$ exists for all $t \in \mathbb{R}$.

Proof: We first prove $E(X^{2m})$ is finite , $\forall m \in \mathbb{N}$, where X has the density $\psi(\cdot)$. This is trivially true for $m = 0$. Assume this is true for $m = k$. Now for any $N \in (0, \infty)$,

$$\begin{aligned}
\int_{-\infty}^{\infty} x^{2k}\psi(x)dx &\geq \int_{-N}^{N} x^{2k}\psi(x)dx \\
&= \frac{1}{2k+1}[\psi(x)x^{2k+1}]_{-N}^{N} - \frac{1}{2k+1}\int_{-N}^{N} \psi'(x)x^{2k+1}dx \\
&= \frac{1}{2k+1}b_{k,N} - \frac{1}{2k+1}\int_{-N}^{N}\Big(x\frac{\psi'(x)}{\psi(x)}\Big)x^{2k}\psi(x)dx,
\end{aligned}$$

where $b_{k,N} = N^{2k+1}(\psi(N) + \psi(-N))$. This implies

$$\begin{aligned}
(2k+1)\int_{-\infty}^{\infty} x^{2k}\psi(x)dx &\geq b_{k,N} - \int_{-N}^{N}\Big(-ax^2 + b\Big)x^{2k}\psi(x)dx \\
&= b_{k,N} + a\int_{-N}^{N} x^{2k+2}\psi(x)dx - b\int_{-\infty}^{\infty} x^{2k}\psi(x)dx \\
\Rightarrow (2k+1+b)\int_{-\infty}^{\infty} x^{2k}\psi(x)dx &\geq b_{k,N} + a\int_{-N}^{N} x^{2k+2}\psi(x)dx,
\end{aligned}$$

since $\int\limits_{-\infty}^{\infty} x^{2k}\psi(x)dx \geq \int\limits_{-N}^{N} x^{2k}\psi(x)dx$, for any $N > 0$. Also, $b_{k,N} = N^{2k+1}(\psi(N) + \psi(-N)) \geq 0$, for any N. This implies that,

$$\begin{aligned}
a\int\limits_{-N}^{N} x^{2k+2}\psi(x)dx &\leq (2k+1+b)\int\limits_{-\infty}^{\infty} x^{2k}\psi(x)dx \\
\Rightarrow \int\limits_{-\infty}^{\infty} x^{2k+2}\psi(x)dx &\leq \frac{2k+1+b}{a}\int\limits_{-\infty}^{\infty} x^{2k}\psi(x)dx \\
\Rightarrow E(X^{2k+2}) &\leq \frac{2k+1+b}{a}E(X^{2k}),
\end{aligned} \tag{5.2.5}$$

where the RHS of (5.2.5) exists by virtue of the induction hypothesis. This implies that all moments of X exist. Iterating (5.2.5) over k we have

$$\begin{aligned}
E(X^{2k+2}) &\leq \frac{(2k+1+b)(2k-1+b)\dots(1+b)}{a^{k+1}} \\
&= \frac{(2k+1)(2k-1)\dots 3.1(1+\frac{b}{2k+1})(1+\frac{b}{2k-1})\dots(1+b)}{a^{k+1}} \\
&\leq \frac{(2k+2)!}{2^{k+1}(k+1)!}(\frac{1+b}{a})^{k+1} \\
\Rightarrow E(X^{2k}) &\leq \frac{(2k)!}{2^k k!}(\frac{1+b}{a})^k,
\end{aligned} \tag{5.2.6}$$

for every $k \in \mathbb{N}$.

Now we find a bound of the absolute odd moments. For any $k \in \mathbb{N}$,

$$\int\limits_{-\infty}^{\infty} |x|^{2k+1}\psi(x)dx \geq \int\limits_{-N}^{N} |x|^{2k+1}\psi(x)dx = \int\limits_{-N}^{0} |x|^{2k+1}\psi(x)dx + \int\limits_{0}^{N} |x|^{2k+1}\psi(x)dx.$$

Consider $\int\limits_{-N}^{0} |x|^{2k+1}\psi(x)dx$. This is equal to

$$\begin{aligned}
\int\limits_{-N}^{0} (-x)^{2k+1}\psi(x)dx &= -[\frac{x^{2k+2}}{2k+2}\psi(x)]_{-N}^{0} + \int\limits_{-N}^{0} \frac{(-x)^{2k+2}}{2k+2}\psi'(x)dx, \\
&= [\frac{(-x)^{2k+1}x\psi(x)}{2k+2}]_{-N}^{0} - \int_{-N}^{0} (-x)^{2k+1}\frac{x\psi'(x)}{\psi(x)}\psi(x)dx \text{ and,} \\
\int\limits_{0}^{N} x^{2k+1}\psi(x)dx &= [\frac{x^{2k+2}}{2k+2}\psi(x)]_{0}^{N} - \int\limits_{0}^{N} \frac{x^{2k+2}}{2k+2}\psi'(x)dx \\
&= \frac{x^{2k+2}}{2k+2}\psi(x)|_{0}^{N} - \int_{0}^{N} \frac{x^{2k+1}}{2k+2}\frac{x\psi'(x)}{\psi(x)}\psi(x)dx. \qquad (5.2.7)
\end{aligned}$$

This implies

$$\begin{aligned}
\int\limits_{-N}^{N} |x|^{2k+1}\psi(x)dx &= [\frac{1}{2k+2}|x|^{2k+1}x\psi(x)]_{-N}^{N} \\
&- \frac{1}{2k+2}\int\limits_{-N}^{N} |x|^{2k+1}x\frac{\psi'(x)}{\psi(x)}\psi(x)dx \\
\Rightarrow (2k+2)\int\limits_{-N}^{N} |x|^{2k+1}\psi(x)dx &\geq c_{k,N} + a\int\limits_{-N}^{N} |x|^{2k+3}\psi(x)dx - b\int\limits_{-N}^{N} |x|^{2k+1}\psi(x)dx,
\end{aligned}$$

where $c_{k,N} = N^{2k+1}(N\psi(N) + N\psi(-N)) \geq 0$. Repeating arguments similar to the even moments of X we have that

$$\begin{aligned}
\int\limits_{-\infty}^{\infty} |x|^{2k+3}\psi(x)dx &\leq \frac{2k+2+b}{a}\int\limits_{-\infty}^{\infty} |x|^{2k+1}\psi(x)ds \\
\Rightarrow E(|X|^{2k+3}) &\leq \frac{2k+2+b}{a}E(|X|^{2k+1}). \qquad (5.2.8)
\end{aligned}$$

Iterating (5.2.8) over k we have

$$
\begin{aligned}
E(|X|^{2k+3}) &\leq \frac{(2k+2+b)(2k+b)\dots(2+b)}{a^{k+1}}E(|X|)\\
&= (2k+2)(2k)\dots 2\frac{(1+\frac{b}{2k+2})(1+\frac{b}{2k})\dots(1+\frac{b}{2})}{a^{k+1}}E(|X|)\\
&\leq 2^{k+1}(k+1)!(\frac{1+b}{a})^{k+1}E(|X|)\\
&\leq \frac{(2k+3)!}{2^{k+1}(k+1)!}(\frac{1+b}{a})^{k+1}E(|X|)\\
\Rightarrow E(|X|^{2k+1}) &\leq 2^k k!(\frac{1+b}{a})^k E(|X|)\\
&\leq \frac{(2k+1)!}{2^k k!}(\frac{1+b}{a})^k E(|X|), \qquad (5.2.9)
\end{aligned}
$$

for $k \in \mathbb{N}$ and the RHS of (5.2.9) since all moments of X exist. Therefore from (5.2.6) and (5.2.9) we have the m.g.f of X as

$$
\begin{aligned}
M_X(t) = E(e^{tX}) &\leq E(e^{t|X|}) = \sum_{n=0}^{\infty}\frac{t^n E(|X|^n)}{n!}\\
&\leq \sum_{n=0}^{\infty}\frac{1}{n!}\Big(\frac{t^2(1+b)}{2a}\Big)^n\Big(tE(|X|)+1\Big)\\
&\leq e^{\frac{t^2}{2}\frac{1+b}{2a}}(e^{tE(|X|)})\\
&= e^{tE(|X|)+\frac{t^2}{2}\frac{1+b}{2a}} \qquad (5.2.10)
\end{aligned}
$$

which exists for all $t \in \mathbb{R}$. This proves the lemma. ■

The next three lemmas deal with condition (5.2.3).

LEMMA 21. If the density $\psi(x)$ satisfies the condition (5.2.3), i.e.,

$$x\frac{\psi'(x)}{\psi(x)} \leq -a_2|x|^\alpha + b_2,$$

for some $\alpha \in (0,2), a_2 > 0,\ b_2 \geq 0$, then all integer moments of $|X|^\alpha$ is finite .

Proof: From the hypothesis we have that

$$\begin{aligned}
\int_{-\infty}^{\infty} |x|^{k\alpha}\psi(x)dx &\geq \int_{-N}^{N} |x|^{k\alpha}\psi(x)dx \\
&= \frac{1}{k\alpha+1}[|x|^{k\alpha}x\psi(x)]_{-N}^{N} - \frac{1}{k\alpha+1}\int_{-N}^{N} |x|^{k\alpha}x\psi'(x)dx,
\end{aligned}$$

using the indefinite integral of $|x|^{\beta}$. Therefore

$$\begin{aligned}
\int_{-\infty}^{\infty} |x|^{k\alpha}\psi(x)dx &= c_{k,N} - \frac{1}{k\alpha+1}\int_{-N}^{N} |x|^{k\alpha}\Big(x\frac{\psi'(x)}{\psi(x)}\Big)\psi(x)dx \\
&\geq c_{k,N} - \frac{a_2}{k\alpha+1}\int_{-N}^{N} |x|^{k\alpha}\psi(x)dx - \frac{b_2}{k\alpha+1}\int_{-N}^{N} |x|^{k\alpha}\psi(x)dx \\
\Rightarrow a_2\int_{-\infty}^{\infty} |x|^{k+1\alpha}\psi(x)ds &\leq (k\alpha+1+b_2)\int_{-\infty}^{\infty} |x|^{k\alpha}\psi(x)dx \\
\Rightarrow E(|X|^{k+1\alpha}) &\leq \frac{k\alpha+1+b_2}{a_2}E(|X|^{k\alpha}),
\end{aligned} \tag{5.2.11}$$

where $c_{k,N} = \frac{1}{k\alpha+1}N^{k\alpha+1}(\psi(N)+\psi(-N)) \geq 0$. Hence by iteration we get that all integer moments of $|X|^{\alpha}$ is finite. ■

The next two lemmas gives the condition when the m.g.f of $\psi(\cdot)$ exists

LEMMA 22. If the density $\psi(x)$ satisfies Equation (5.2.3) for $\alpha \in [1,2)$ then the m.g.f of $\psi(\cdot)$ exists in a neighbourhood of zero.

Proof: From Equation (5.2.11) we have

$$\begin{aligned}
E(|X|^{k\alpha}) &\leq \frac{(k-1)\alpha+1+b_2}{a_2}E(|X|^{(k-1)\alpha}) \\
&\vdots \\
&\leq \Big(\frac{\overline{k-1}\alpha+1+b_2}{a_2}\Big)\Big(\frac{\overline{k-2}\alpha+1+b_2}{a_2}\Big)\dots\Big(\frac{1+b_2}{a_2}\Big)
\end{aligned} \tag{5.2.12}$$

Applying Jensen's inequality to the above equation we get

$$\begin{aligned}
E(|X|^k) &\leq E(|X|^{k\alpha})^{\frac{1}{\alpha}} \\
&\leq \frac{1}{a_2^{\frac{k}{\alpha}}}\Big((\overline{k-1}\alpha+1+b_2)(\overline{k-2}\alpha+1+b_2)\ldots(1+b_2)\Big)^{\frac{1}{\alpha}} \\
&\leq \frac{1}{a_2^{\frac{k}{\alpha}}}(\overline{k-1}\alpha+1+b_2)(\overline{k-2}\alpha+1+b_2)\ldots(1+b_2),
\end{aligned}$$

since $b_2 \geq 0, \alpha \geq 1$. Therefore

$$E(|X|^k) \leq \frac{\alpha^k}{a_2^{\frac{k}{\alpha}}}(\overline{k-1}+\frac{1+b_2}{\alpha})(\overline{k-2}+\frac{1+b_2}{\alpha})\ldots(\frac{1+b_2}{\alpha}).$$

Since all the absolute moments exist we have $M_X(t) =$

$$\begin{aligned}
E(e^{tX}) &\leq E(e^{t|X|}) \\
&= \sum_{k=0}^{\infty}\frac{t^k}{k!}\frac{\alpha^k}{a_2^{\frac{k}{\alpha}}}(\overline{k-1}+\frac{1+b_2}{\alpha})(\overline{k-2}+\frac{1+b_2}{\alpha})\ldots(\frac{1+b_2}{\alpha}) \\
&= \sum_{k=0}^{\infty}(-1)^k\Big(\frac{t\alpha}{a_2^{\frac{1}{\alpha}}}\Big)^k\frac{(-\frac{1+b_2}{\alpha}-\overline{k-1})(-\frac{1+b_2}{\alpha}-\overline{k-2})\ldots(-\frac{1+b_2}{\alpha})}{k!}.
\end{aligned}$$

The infinite sum on the RHS of the above equation is finite for $|t| < \frac{a_2^{\frac{1}{\alpha}}}{\alpha}$ and the value is $(1-\frac{t\alpha}{a_2^{\frac{1}{\alpha}}})^{-\frac{1+b_2}{\alpha}}$. This proves the lemma. ■

The next lemma is for $\alpha \in (0,1)$

LEMMA 23. If the density satisfies the condition in Equation (5.2.3) for $\alpha \in (0,1)$ then the m.g.f does not exist in any neighborhood of zero.

Proof: We show that $E(e^{tX})$ is not finite for any $t < 0$. Now for $x < 0$, Equation (5.2.3) is

$$\begin{aligned}
\frac{\psi'(x)}{\psi(x)} &\geq -a_2\frac{(-x)^\alpha}{x} + \frac{b_2}{x} \\
&= a_2(-x)^{\alpha-1} + \frac{b_2}{x} \\
&\leq -a_2\frac{-x^\alpha}{x} + \frac{b_2}{x} \qquad (5.2.13) \\
\Rightarrow \log(\psi(x)) &\geq a_2\int(-x)^{\alpha-1}dx + b_2\log(|x|) + c \\
\Rightarrow \psi(x) &\geq K|x|^{b_2}e^{-\frac{a_2}{\alpha}|x|^\alpha} \\
\Rightarrow \int_{-\infty}^{0} e^{tx}\psi(x)dx &\geq K\int_{-\infty}^{0}|x|^{b_2}e^{tx-\frac{a_2}{\alpha}|x|^\alpha}dx \qquad (5.2.14)
\end{aligned}$$

However the integral on the RHS of the above equation diverges to ∞ for any $t < 0$, since $\alpha < 1$. Hence the m.g.f of X does not exist in a neighbourhood of zero. This proves the lemma.

■

REMARK 27. Although the m.g.f does not exist it is clear from Equation (5.2.12) that all moments exist for $\alpha \in (0, 1)$ since $\forall\ n \geq 1\ \exists K$ such that $n < K\alpha \Rightarrow E(|X|^n) \leq E(|X|^{k\alpha})^{\frac{n}{k\alpha}} < \infty$.

The next lemma deals with the condition in (5.2.4).

LEMMA 24. If the density satisfy condition in Equation (5.2.4), i.e,

$$x\frac{\psi'(x)}{\psi(x)} \leq -a_3,$$

for some $a_3 > 0$, then $E(X^{2k}) = \infty$, $\forall k \geq k_0$, for some $k_0 \in \mathbb{N}$.

Proof: For $x < 0$

$$\begin{aligned}
\frac{\psi'(x)}{\psi(x)} &\geq -\frac{a_3}{x} \\
\Rightarrow \log\psi(x) &\geq -a_3\log|x| + c \\
\Rightarrow \psi(x) &\geq K_0|x|^{-a_3} \\
\Rightarrow \int_{-\infty}^{0}|x|^k\psi(x)dx &= \infty,
\end{aligned}$$

for all $a_3 > 0$ and $k > a_3 - 1$. Consequently all moments of X does not exist. ■

Combining the above lemmas we have the theorem:

THEOREM 6. If the density $\psi(\cdot)$ satisfy condition (5.2.3) with $\alpha \in [1,2)$ then the m.g.f exists in a neighbourhood of zero. If it satisfies (5.2.3) with $\alpha \in (0,1)$ then the m.g.f does not exist in any neighbourhood of zero, but all the moments exists. If the density satisfies (5.2.4) then all but finitely many moments are infinite.

5.2.2 Light tailed target distribution

In this section the proposal distribution is fixed to be standard Normal and the target density $\psi(\cdot)$ satisfies condition in (5.2.2). Recall, the SDE of the limiting AMCMC is given by (Equation 3.4.1),

$$dX_t = \frac{\theta_t^2}{2}\frac{\psi'(X_t)}{\psi(X_t)}dt + \theta_t dW_t \tag{5.2.15}$$

$$d\theta_t = \theta_t\Big(q - \frac{\theta_t}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}\Big)dt \tag{5.2.16}$$

where $\psi(\cdot)$ is the target distribution. We first prove two lemmas

LEMMA 25. If the process $\{X_t, \theta_t\}$ satisfies Equation (5.2.15) and (5.2.16) and $\eta_t = \frac{1}{\theta_t}$, where $\frac{\psi'(x)}{\psi(x)}$ satisfies the growth condition given in Equation (5.2.1), i.e.,

$$\frac{|\psi'(x)|}{\psi(x)} \le a|x| + b,\ \ a \ge 0, b \ge 0,$$

and the condition in Equation (5.2.2) and Equation (5.2.3) then the moments of η_t^{2k} are uniformly bounded in $t > 0$, for all $k \in \mathbb{N}$, i.e.,

$$\sup_{t>1} E(\eta_t^{2k}) < \infty,$$

if $E(\eta_0^{2k}) < \infty$ and $E(X_0^{2k}) < \infty$.

Proof: From Equation (5.2.16) the SDE corresponding to η_t is

$$d\eta_t = -q\eta_t dt + \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}dt.$$

From Equation (5.2.17) we have

$$\begin{aligned}
\Rightarrow d\eta_t + q\eta_t dt &\leq \frac{1}{\sqrt{2\pi}}(a|X_t| + b) \\
\Rightarrow d(e^{qt}\eta_t) &\leq a\frac{1}{\sqrt{2\pi}}e^{qt}|X_t|dt + be^{qt}dt \\
\Rightarrow \eta_t &\leq \eta_0 e^{-qt} + \frac{1}{\sqrt{2\pi}}(a\int_0^t e^{-q(t-s)}|X_s|ds + b\int_0^t e^{-q(t-s)}ds) \\
\Rightarrow E(\eta_t^{2k}) &\leq C_0\Big(e^{-2kqt}E(\eta_0^{2k}) + \frac{1}{(2\pi)^k}\Big(a^{2k}E(\int_0^t e^{-q(t-s)}|X_s|ds)^{2k} + (\frac{b}{q})^{2k}(1-e^{-qt})^{2k}\Big)\Big),
\end{aligned} \tag{5.2.17}$$

for some constant $C_0 \in (0, \infty)$. Now using the same techniques as in the proof of Lemma 15 we have

$$\begin{aligned}
E\Big(e^{-qt}\int_0^t e^{qs}|X_s|\Big)^{2k} &= E\Big(\frac{1-e^{-qt}}{q}\frac{q}{e^{qt}-1}\int_0^t e^{qu}|X_u|du\Big)^{2k} \\
&= \Big(\frac{1-e^{-qt}}{q}\Big)^{2k}E\Big(\frac{q}{e^{qt}-1}\int_0^t e^{qu}|X_u|du\Big)^{2k} \\
&\leq \Big(\frac{1-e^{-qt}}{q}\Big)^{2k}\Big(\frac{q}{e^{qt}-1}\int_0^t e^{qu}E|X_u|^{2k}du\Big) \\
&\leq M\Big(\frac{1-e^{-qt}}{q}\Big)^{2k},
\end{aligned} \tag{5.2.18}$$

since it is proved in Theorem 7 and Theorem 8 that $\sup_{t>0} E(X_t^{2k}) < M < \infty$ for some M depending on k under the condition of Equation (5.2.2) and Equation (5.2.3). Consequently,

$$\sup_{t>0} E(\eta_t^{2k}) < \infty.$$

■

Remark 28. From the bounds of $E(\eta_t^{2k})$ given in Equation (5.2.17) and Equation (5.2.18) we have that for all $k \in \mathbb{N}$

$$\sup_{t>0} E(\eta_t^k) \quad \leq \quad M_1^k,$$

for some constant $M_1 (= M_1(k))$ which depends on $E(\eta_0^k)$ and $E(X_0^k)$. Hence by the hypothesis of Lemma 25 $M_1 < \infty$ for all $k \in \mathbb{N}$. In addition if we assume that the m.g.f of X_0 and η_0 exist for each $t \in \mathbb{R}$ then

$$\sup_{t>0} E(e^{s\eta_t}) \quad < \quad \infty,$$

for all $s \in \mathbb{R}$.

Lemma 26. If θ_t satisfies the SDE given in Equation (5.2.16) then $\lim_{t\to\infty} E(e^{-r\int_0^t \theta_s^2 ds}) = 0$ for every $r > 0$.

Proof: The proof is similar to the proof of Lemma 19. We prove for $r = 1$. The result is true for any $r > 0$. Define $F_1(t) = \int_0^t \theta_s^2 ds$. Then

$$\frac{F_1(t)}{t} \quad = \quad \frac{1}{t}\int_0^t \theta_s^2 ds \geq \frac{1}{\frac{1}{t}\int_0^t \frac{1}{\theta_s^2} ds} = \frac{1}{\frac{1}{t}\int_0^t \eta_s^2 ds} \tag{5.2.19}$$

where the last but one inequality follows from Jensen's (by taking $f(x) = \frac{1}{x}$, $x > 0$ which is convex). This implies

$$\frac{1}{\frac{1}{t}F_1(t)} \quad \leq \quad \frac{1}{t}\int_0^t \eta_s^2 ds, \Rightarrow \frac{1}{F_1(t)} \leq \frac{1}{t}\frac{1}{t}\int_0^t \eta_s^2 ds. \tag{5.2.20}$$

Therefore,

$$
\begin{aligned}
e^{-F_1(t)} &= \frac{1}{e^{F_1(t)}} \le \frac{1}{F_1(t)} \text{ (since } e^x \ge x,\ \forall x > 0) \\
&\le \frac{1}{t}\frac{1}{t}\int_0^t \eta_s^2 ds, \text{ from (5.2.20)} \\
\Rightarrow E(e^{-F_1(t)}) &\le \frac{1}{t}E\Big(\frac{1}{t}\int_0^t \eta_s^2 ds\Big) \le \frac{1}{t}C.
\end{aligned}
$$

where $C = \sup_{t>0} E(\frac{1}{t}\int_0^t \eta_s^2 ds) < \infty$, from Lemma 25. So

$$
\lim_{t\to\infty} E(e^{-F_1(t)}) = 0.
$$

■

For densities of the form Equation (5.2.2) the m.g.f of the limiting distribution exists for all $t \in \mathbb{R}$.

THEOREM 7. If the density of the target distribution satisfies the condition in Equation (5.2.2), i.e.,

$$
x\frac{\psi'(x)}{\psi(x)} \le -a|x|^2 + b,
$$

for some $a > 0,\ b \ge 0$ then the m.g.f. of the limiting distribution of X_t exists for all $s \in \mathbb{R}$ if the m.g.f of X_0 exists for all $s \in \mathbb{R}$.

Proof: Fix any $k \in \mathbb{N}$. Applying Itô's lemma to X_t^{2k} we get :

$$
\begin{aligned}
dX_t^{2k} &= 2kX_t^{2k-1}dX_t + \frac{1}{2}2k(2k-1)X_t^{2k-2}\theta_t^2 dt \\
&= \Big(2kX_t^{2k-1}\frac{\theta_t^2}{2}\frac{\psi'(X_t)}{\psi(X_t)} + k(2k-1)X_t^{2k-2}\theta_t^2\Big)dt + 2kX_t^{2k-1}\theta_t dW_t \\
&\le k\theta_t^2\Big(-aX_t^{2k} + bX_t^{2k-2} + (2k-1)X_t^{2k-2}\Big)dt + 2kX_t^{2k-1}\theta_t dW_t \\
&= -ak\theta_t^2 X_t^{2k}dt + \theta_t^2 X_t^{2k-2}\Big(kb + k(2k-1)\Big)dt \\
&+ 2k\theta_t X_t^{2k-1}dW_t, \qquad (5.2.21)
\end{aligned}
$$

using the hypothesis (5.2.2). Also for any positive real c_k there exists large enough d_k so that

$$x^{2k-2} < c_k x^{2k} + d_k \ \forall x \in \mathbb{R}.$$

Applying this inequality to (5.2.21) we get

$$\begin{aligned} dX_t^{2k} &\leq -ak\theta_t^2 X_t^{2k}dt + \Big(kb + k(2k-1)\Big)\theta_t^2(c_k X_t^{2k} + d_k)dt + 2kX_t^{2k-1}dW_t \\ &= \Big(-ak + kbc_k + k(2k-1)c_k\Big)\theta_t^2 X_t^{2k}dt + d_k\Big(kb + k(2k-1)\Big)\theta_t^2 dt \\ &+ 2kX_t^{2k-1}dW_t. \end{aligned} \tag{5.2.22}$$

Define $A_k := -\Big(-ak + kbc_k + k(2k-1)c_k\Big)$ where we choose c_k small so that $A_k > 0$. That is,

$$\begin{aligned} ak - kbc_k - k(2k-1)c_k &> 0 \\ \Rightarrow c_k &< \frac{a}{2k-1+b}. \end{aligned}$$

To find a bound for d_k, define $f(x) = c_k x^{2k} - x^{2k-2} + d_k$. Then

$$\begin{aligned} f'(x) &= 2kc_k x^{2k-1} - (2k-2)x^{2k-3} \\ &= 2kx^{2k-3}(c_k x^2 - \frac{k-1}{k}). \end{aligned}$$

Hence $x = \pm\sqrt{\frac{k-1}{kc_k}}$ is local minima and $x = 0$ is the local maxima. Therefore

$$\begin{aligned} f(\sqrt{\frac{k-1}{kc_k}}) &\geq 0 \\ \Rightarrow c_k(\frac{k-1}{kc_k})^k - (\frac{k-1}{kc_k})^{k-1} + d_k &\geq 0 \\ \Rightarrow (\frac{k-1}{kc_k})^{k-1}(c_k\frac{k-1}{kc_k} - 1) + d_k &\geq 0 \\ \Rightarrow d_k &\geq (\frac{k-1}{kc_k})^{k-1}\frac{1}{k}. \end{aligned} \tag{5.2.23}$$

Choose $c_k = \frac{a}{2k+b} < \frac{a}{2k+b-1}$ and $d_k = (\frac{2k+b}{a})^{k-1}\frac{1}{k}$. Then $A_k = \frac{ak}{2k+b}$. Also define $G_k := d_k\Big(kb + k(2k-1)\Big)$. We then have

$$dX_t^{2k} + A_k\theta_t^2 X_t^{2k} \leq G_k\theta_t^2 dt + 2k\theta_t X_t^{2k-1}dW_t.$$

Multiplying by the integrating factor $e^{\frac{ak}{2k+b}\int\limits_0^t \theta_s^2 ds} = e^{A_k\int\limits_0^t \theta_s^2 ds} := e^{M_t}$ on both sides we get

$$\begin{aligned}
d\Big(X_t^{2k}e^{M_t}\Big) &\leq e^{M_t}\Big(G_k\theta_t^2 dt + 2k\theta_t X_t^{2k-1}dW_t\Big) \\
\Rightarrow X_t^{2k}e^{M_t} &\leq X_0^{2k} + G_k\int\limits_0^t e^{M_s}\theta_s^2 ds + 2k\int\limits_0^t e^{M_s}X_s^{2k-1}\theta_s dW_s \qquad (5.2.24) \\
\Rightarrow E(X_t^{2k}) &\leq E(X_0^{2k}e^{-M_t}) + G_k E\Big(e^{-M_t}\int\limits_0^t e^{M_s}\theta_s^2 ds\Big) \\
&+ 2kE\Big(e^{-M_t}\int\limits_0^t e^{M_s}X_s^{2k-1}\theta_s dW_s\Big) \\
&= E(X_0^{2k}e^{-M_t}) + \frac{G_k}{A_k}E\Big(e^{-M_t}\Big(\int\limits_0^t d(e^{M_s})\Big)\Big) \\
&= E(X_0^{2k})E(e^{-M_t}) + \frac{G_k}{A_k}\Big(1 - E(e^{-M_t})\Big), \qquad (5.2.25)
\end{aligned}$$

arguing as in the proof of Lemma 12 in Chapter 4 that $E(e^{-M_t}\int\limits_0^t e^{M_s}X_s^{2k-1}\theta_s dW_s) = 0$(it is a symmetric random variable whose second moment is finite.) Since $M_t \geq 0, E(X_t^{2k})$ is uniformly bounded in $t > 0$ from Equation (5.2.25).

Now $\frac{G_k}{A_k} = \frac{kd_k(2k-1+b)}{A_k} = (\frac{2k+b}{a})^k(\frac{2k+b-1}{k})$. By Lemma 26 $\lim\limits_{t\to\infty} E(e^{-M_t}) = 0$ which implies, by DCT,

$$\begin{aligned}
\lim_{t\to\infty} E(X_t^{2k}) &\leq \frac{G_k}{A_k} = (\frac{(2k+b)}{a})^k(\frac{2k+b-1}{k}) \leq \frac{a}{k}(\frac{2k+b}{a})^{k+1} \\
&= a(\frac{(2+\frac{b}{k})}{a})^{k+1}k^k \leq a(\frac{2+b}{a})^{k+1}k^k \\
&= (2+b)(\sqrt{\frac{2+b}{a}})^{2k}k^{\frac{2k}{2}}. \qquad (5.2.26)
\end{aligned}$$

Therefore

$$\begin{aligned}
\lim_{t\to\infty} E|X_t|^{(2k-1)} &\leq \lim_{t\to\infty} E(X_t^{2k})^{\frac{2k-1}{2k}} \\
&\leq (a(\frac{2+b}{a})^{k+1}k^k)^{\frac{2k-1}{2k}} \\
&= a^{\frac{2k-1}{2k}}(\frac{2+b}{a})^{\frac{2k-1}{2}}(\frac{2+b}{a})^{\frac{2k-1}{2k}}k^{\frac{2k-1}{2}} \\
&= (2+b)^{\frac{2k-1}{2k}}(\frac{2+b}{a})^{\frac{2k-1}{2}}k^{\frac{2k-1}{2}} \\
&\leq (2+b)(\sqrt{\frac{2+b}{a}})^{2k-1}k^{\frac{2k-1}{2}} \qquad (5.2.27)
\end{aligned}$$

From (5.2.26) and (5.2.27) we have

$$\frac{t^k}{k!}E|X|^k \leq (2+b)\frac{\left(t\sqrt{\frac{2+b}{a}}\right)^k}{k!}k^{\frac{k}{2}}, \qquad (5.2.28)$$

for $k \geq 1$. Therefore

$$M_X(t) \leq E(e^{t|X|}) \leq 1+(2+b)\sum_{k=1}^{\infty}\frac{(t\sqrt{\frac{2+b}{a}})^k k^{\frac{k}{2}}}{k!}.$$

The k^{th} term of the above power series on the right is given by $a_k t^k$, where a_k is equal to

$$(2+b)\frac{(t\sqrt{\frac{2+b}{a}})^k}{k!}k^{\frac{k}{2}}.$$

Using Stirling's approximation we have

$$a_k \approx \frac{2+b}{\sqrt{2\pi}}\frac{(e\sqrt{\frac{2+b}{a}})^k}{k^{\frac{k}{2}+\frac{1}{2}}}.$$

Therefore

$$\Rightarrow \limsup_{k\to\infty}|a_k|^{\frac{1}{k}} = \limsup_{k\to\infty}(\frac{2+b}{\sqrt{2\pi}})^{\frac{1}{k}}\frac{e\sqrt{\frac{2+b}{a}}}{k^{\frac{1}{2}+\frac{1}{2k}}} = 0$$

which implies the radius of convergence of this power series is

$$R = \frac{1}{\limsup\limits_{k\to\infty} |a_k|^{\frac{1}{k}}} = \infty.$$

Hence $M_X(t) < \infty$ for every $t \in \mathbb{R}$. This proves the Theorem.

$\blacksquare$

REMARK 29. Since the second moments of $(X_t, \eta_t)_{t>0}$ are uniformly bounded the process is tight.

For densities satisfying condition (5.2.3) for $\alpha \in [1, 2)$ we first prove a lemma.

LEMMA 27.

$$\lim_{t\to\infty} E(e^{-k\int\limits_0^t \frac{\theta_s^2}{|X_s|+b}ds}) = 0,$$

for any $k > 0$.

Proof: The proof is similar to the proof of Lemma 26. We prove only for $k = 1$. The result is true for any $k > 0$. Define, $F_1(t) = \int\limits_0^t \frac{\theta_s^2}{|X_s|+b}ds$

$$\frac{F_1(t)}{t} = \frac{1}{t}\int\limits_0^t \frac{\theta_s^2}{|X_s|+b}ds \geq \frac{1}{\frac{1}{t}\int\limits_0^t \frac{|X_t|+b}{\theta_s^2}ds} = \frac{1}{\frac{1}{t}\int\limits_0^t (|X_s|+b)\eta_s^2 ds} \tag{5.2.29}$$

where the last but one inequality follows from Jensen's (by taking $f(x) = \frac{1}{x}$, $x > 0$ which is convex). This implies

$$\frac{1}{\frac{1}{t}F_1(t)} \leq \frac{1}{t}\int\limits_0^t (|X_s|+b)\eta_s^2 ds \Rightarrow \frac{1}{F_1(t)} \leq \frac{1}{t}\frac{1}{t}\int\limits_0^t (|X_s|+b)\eta_s^2 ds. \tag{5.2.30}$$

Therefore,

$$\begin{aligned}
e^{-F_1(t)} &= \frac{1}{e^{F_1(t)}} \leq \frac{1}{F_1(t)} \text{ (since } e^x \geq x, \ \forall x > 0) \\
&\leq \frac{1}{t}\frac{1}{t}\int_0^t (|X_s| + b)\eta_s^2 ds \\
\Rightarrow E(e^{-F_1(t)}) &\leq \frac{1}{t}E\Big(\frac{1}{t}\int_0^t (|X_s| + b)\eta_s^2 ds\Big) \\
&= \frac{1}{t}\Big(\frac{1}{t}\int_0^t E(|X_s| + b)\eta_s^2 ds\Big) \\
&\leq \frac{1}{t}\Big(\frac{1}{t}\int_0^t \sqrt{E(|X_s| + b)^2 E(\eta_s^4) ds}\Big)
\end{aligned}$$

From the next Theorem we have that the moments of X_t are uniformly bounded. From Lemma 25 the moments of η_t are also uniformly bounded (under the hypothesis that densities satisfy (5.2.3) with $\alpha \in [1, 2)$). Consequently

$$\lim_{t\to\infty} E(e^{-F_1(t)}) = 0. \tag{5.2.31}$$

■

THEOREM 8. If the density $\psi(\cdot)$ satisfy condition (5.2.3) for $\alpha \in [1, 2)$, i.e.,

$$x\frac{\psi'(x)}{\psi(x)} \leq -a_2|x|^\alpha + b_2, \ a_2 > 0, b_2 \geq 0,$$

then the m.g.f of the limiting distribution of X_t exists in a neighborhood of zero whenever m.g.f of X_0 exists

Proof: Applying Itô's lemma to X_t^{2k} and using Equation (5.2.3) we have

$$dX_t^{2k} \leq k\theta_t^2 X_t^{2k-2}\Big(-a_2|X_t|^\alpha + b_2 + 2k - 1\Big)dt + 2kX_t^{2k-1}\theta_t dW_t. \tag{5.2.32}$$

Now for any $\alpha \in (1, 2)$ we have that

$$|X_t| \leq a_2|X_t|^\alpha + b_a,$$

for some $a_2 > 0$ and $b_a > 0$ chosen sufficiently large depending on a_2. This implies that

$$\begin{aligned} -a_2|X_t|^\alpha &\leq -|X_t| + b_a \\ \Rightarrow -a_2|X_t|^\alpha + b_2 + 2k - 1 &\leq -|X_t| + b_3, \end{aligned}$$

where $b_3 = b_a + b_2 + 2k - 1$. Substituting in Equation (5.2.32) we have

$$\begin{aligned} dX_t^{2k} &\leq k\theta_t^2 X_t^{2k-2}\Big(-|X_t| + b_3\Big) + 2kX_t^{2k-1}\theta_t dW_t \qquad (5.2.33) \\ &= -k\theta_t^2 X_t^{2k-2}\Big(|X_t| - b_3\Big) + 2kX_t^{2k-1}\theta_t dW_t \\ &= -k\theta_t^2 X_t^{2k-2}\Big(\frac{X_t^2 - b_3^2}{|X_t| + b_3}\Big) + 2kX_t^{2k-1}\theta_t dW_t \\ \Rightarrow dX_t^{2k} + k\theta_t^2\frac{X_t^{2k}}{|X_t| + b_3} &= b_3^2 k\theta_t^2\frac{X_t^{2k-2}}{|X_t| + b_3} + 2kX_t^{2k-1}\theta_t dW_t \\ &\leq b_3^2 k\theta_t^2\frac{aX_t^{2k} + b}{|X_t| + b_3} + 2kX_t^{2k-1}\theta_t dW_t. \end{aligned}$$

Using the inequality

$$x^{2k-2} \leq ax^{2k} + b,$$

for small $a > 0$ and large $b \geq 0$ (where both a and b depends on k), we have,

$$\Rightarrow dX_t^{2k} + k(1 - ab_3^2)X_t^{2k}\frac{\theta_t^2}{|X_t| + b_3} \leq kbb_3^2\frac{\theta_t^2}{|X_t| + b_3} + 2kX_t^{2k-1}\theta_t dW_t.$$

Using e^{M_t} as the integrating factor, integrating both sides and taking expectations we have

$$E(X_t^{2k}) \leq E(X_0^{2k})E(e^{-M_t}) + \frac{G_k}{A_k}(1 - e^{-M_t}),$$

where $A_k = k(1 - ab_3^2) > 0, M_t = A_k \int\limits_0^t \frac{\theta_s^2}{|X_s|+b_3} ds$ and $G_k = kbb_3^3$. Arguing as in Lemma 12 the expectation of the Itô Integral $E(e^{-M_t} \int\limits_0^t X_s^{2k-1}\theta_s dW_s)$ is zero. This proves that the moments of X_t are uniformly bounded in $t > 0$.

From Lemma 27 and the definition of M_t we have that

$$\lim_{t\to\infty} E(e^{-M_t}) = 0,$$

and so

$$E(X_t^{2k}) \;\leq\; \frac{G_k}{A_k} = \frac{bb_3^2}{1-ab_3^2}.$$

Choose $a = \frac{1}{2b_3^2}$. Therefore

$$\frac{G_k}{A_k} \;=\; 2bb_3^2 = \frac{b}{a}.$$

Now from Equation (5.2.23) we have that for every $k \in \mathbb{N}$, $b > (\frac{k-1}{ka})^{\overline{k-1}}\frac{1}{k}$. Choose $b = \frac{1}{a^{\overline{k-1}}}$. Therefore

$$\begin{aligned}\limsup_{t\to\infty} E(X_t^{2k}) \leq \frac{G_k}{A_k} \;&=\; \frac{1}{a^k} = 2^k(b_2 + b_a + 2k - 1)^{2k} \\ &\leq\; 2^{\frac{2k}{2}}(b_2 + b_a + 2k)^{2k}.\end{aligned}$$

Now,

$$\begin{aligned}\limsup_{t\to\infty} E(|X_t|^{2k-1}) \;&\leq\; \limsup_{t\to\infty} E(X_t^{2k})^{\frac{2k-1}{2k}} \\ &\leq\; 2^{\frac{2k-1}{2}}(b_2 + b_a + 2k - 1)^{2k-1}.\end{aligned}$$

Therefore

$$\begin{aligned}\limsup_{s\to\infty} \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X_s^k) \;&\leq\; \lim_{s\to\infty} \sum_{k=0}^{\infty} \frac{t^k}{k!} E(|X_s|^k) \\ &=\; \sum_{k=0}^{\infty} (\sqrt{2}t)^k \frac{(b_2 + b_a + k)^k}{k!}.\end{aligned}$$

The coefficient of t^k will be

$$\begin{aligned} a_k &= (\sqrt{2})^k \frac{(b_2 + b_a + k)^k}{k!} \\ \Rightarrow \limsup_{k\to\infty} |a_k|^{\frac{1}{k}} &= \sqrt{2} \limsup_{k\to\infty} \frac{b_2 + b_a + k}{(\sqrt{2\pi})^{\frac{1}{k}} e^{-1} k^{1+\frac{1}{2k}}} = \sqrt{2}e, \end{aligned}$$

using Stirling's approximation. Hence the radius of convergence in $R = \frac{1}{\limsup |a_k|^{\frac{1}{k}}} = \frac{1}{\sqrt{2}e}$. Therefore the m.g.f converges if $t < \frac{1}{\sqrt{2}e}$.
For $\alpha = 1$ we have from Equation (5.2.32)

$$dX_t^{2k} \leq k\theta_t^2 X_t^{2k-2}\Big(-a_2|X_t| + b_2 + 2k - 1\Big)dt + 2kX_t^{2k-1}\theta_t dW_t.$$

which is same as Equation (5.2.33) with some change in constants. Therefore the theorem is true for $\alpha = 1$. This proves the theorem. ■

Lemma 28. **Hypoelliptic condition:** Let the density $\psi(\cdot)$ have finitely many modes and also satisfy the condition $\psi\psi'' - \psi'^2 \neq 0$ for any $x \in \mathbb{R}$. Then The vector fields defined by the SDEs (5.2.1) where the target distribution $\psi(\cdot)$ satisfies Equation (5.2.2) and (5.2.3) with $\alpha \in [1, 2)$ satisfies the Hörmander's hypoelliptic conditions.

Proof: Let $\mathbf{y} = (x, \eta)$. Define,

$$b_\epsilon(\mathbf{y}) = \Big(\frac{1}{2\eta^2}h(x), -q\eta + \frac{1}{\sqrt{2\pi}}g_\epsilon(x)\Big),$$

where for a fixed target distribution $\psi(\cdot)$, $h(x) = \frac{\psi'(x)}{\psi(x)}$ and $g_\epsilon(x)$ is a differentiable function in x such that $\lim_{\epsilon\downarrow 0} g_\epsilon(x) = |h(x)|$ and $\sigma(\mathbf{y})$ is as defined in Section 4.2.2 of Chapter 4. Such a $g_\epsilon(x)$ can be constructed in the manner following Remark 17, viz.,

$$\begin{aligned} g_\epsilon(y) &= \int_{-\infty}^{\infty} |h(x)| \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{1}{2\epsilon^2}(y-x)^2} dx \\ &= \int_{-\infty}^{\infty} |\frac{\psi'(x)}{\psi(x)}| \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{1}{2\epsilon^2}(y-x)^2} dx. \end{aligned}$$

Lemma 29 shows that the family $\{g_\epsilon(\cdot)\}$ is equicontinuous for all $y \in \mathbb{R}$. Note that $|h(x)|$ may not be differentiable at modal points. This will be the drift and the diffusion vectors of the diffusion process $(X_t^\epsilon, \eta_t^\epsilon)$. Therefore the vector fields will be

$$\begin{aligned} A_0^\epsilon(\mathbf{y}) &= \frac{h(x)}{2\eta^2}\frac{\partial}{\partial x} + (-q\eta + \frac{1}{\sqrt{2\pi}} g_\epsilon(x))\frac{\partial}{\partial \eta}, \\ A_1(\mathbf{y}) &= \frac{1}{\eta}\frac{\partial}{\partial x}. \end{aligned}$$

Therefore the vectors corresponding to $A_1(\mathbf{y})$ and $[A_1(\mathbf{y}), A_0^\epsilon(\mathbf{y})]$ will be $\left(\frac{1}{\eta}, 0\right)^T$ and $\left(\frac{1}{\eta^2}\left(\frac{h'(x)}{2\eta} - q\eta + \frac{1}{\sqrt{2\pi}} g_\epsilon(x)\right), \frac{1}{\sqrt{2\pi}}\frac{1}{\eta} g_\epsilon'(x)\right)^T$ respectively. Note that $\theta_t = 1/\eta_t > 0$ almost surely, since by Lemma 25 we have $\sup_{t>0} E(\eta_t^2) < \infty$. Thus, zero set of $\{X_t\}$, has Lebesgue measure zero almost surely since the zero set of $\{W_t\}_{t\geq 0}$ has Lebesgue measure zero. Under the hypothesis of the lemma, $g_\epsilon'(x) \neq 0$ for any $x \in \mathbb{R}$ and consequently the two vector fields span upper half plane of $\mathbb{R}^2$, for $x \neq 0$. This proves the lemma. ■

Lemma 29. If $\frac{|\psi'(\cdot)|}{\psi(\cdot)}$ satisfy the Holder's continuity condition with index $\alpha > 0$, i.e.,

$$|\frac{|\psi'(y_1)|}{\psi(y_1)} - \frac{\psi'(y_2)}{\psi(y_2)}| \leq |y_1 - y_2|^\alpha, y_1, y_2 \in \mathbb{R}$$

then $\{g_\epsilon(\cdot)\}$ is equicontinuous in $\epsilon > 0$ for all $y \in \mathbb{R}$.

Proof: From the definition of $g_\epsilon(\cdot)$ we have

$$\begin{aligned} g_\epsilon(y) &= \int_{-\infty}^{\infty} \frac{|\psi'(x)|}{\psi(x)} \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{1}{2\epsilon^2}(y-x)^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{|\psi'(y+\epsilon z)|}{\psi(y+\epsilon z)} e^{-\frac{1}{2}z^2} dx, \text{ writing } z = \frac{x-y}{\epsilon}, \\ \Rightarrow |g_\epsilon(y_1) - g_\epsilon(y_2)| &\leq \frac{1}{\sqrt{2\pi}} \int_0^t |\frac{\psi'(y_1+\epsilon z)}{\psi(y_1+\epsilon z)} - \frac{\psi'(y_2+\epsilon z)}{\psi(y_2+\epsilon z)}| e^{-\frac{1}{2}z^2} dz \\ &= |y_1 - y_2|^\alpha \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = |y_1 - y_2|^\alpha \\ \Rightarrow \sup_{\epsilon>0} |g_\epsilon(y_1) - g_\epsilon(y_2)| &\leq |y_1 - y_2|^\alpha \end{aligned}$$

This proves the lemma. ■

Remark 30. Under the hypothesis of the above lemma the vector fields corresponding to the SDE of $\{X_t, \eta_t\}$ spans the upper half of $\mathbb{R}^2$, except possible for a finitely many points, and therefore the transition probability of the SDE admits a smooth density. By the finiteness of the moments of X_t and η_t the process is jointly tight and hence the invariant probability exists which admits a density.

5.2.3 Finiteness of time average of moments of θ_t

In this section we show that the time averaged moments of θ is uniformly bounded in $t > 1$, i.e.,

$$\sup_{t>1} \frac{1}{t} \int\limits_0^t E(\theta_s^k) ds < C, \ \forall k \in \mathbb{N},$$

for some constant $C > 0$, for the general target density $\psi(\cdot)$ assuming condition Equation (5.2.1). Here C will stand for a generic finite positive constant that might take different values in different situations. This proof essentially emulates the proof in Section 4.2.1.4. In this section we assume that X_0 and θ_0 admit finite moments of all orders. For non-random initial data this is trivially true. We first prove a lemma that will be used in the Theorem later.

LEMMA 30. If X_t and θ_t are solutions to Equation (5.2.15) and (5.2.16), $E(X_0^k)$ and $E(\theta_0^k)$ is finite $\forall k \in \mathbb{N}$, the target distribution $\psi(\cdot)$ satisfy condition in Equation (5.2.1), viz.,

$$\frac{|\psi(x)|}{\psi(x)} \quad \leq \quad a|x| + b, \ \ a, b \geq 0, \forall x \in \mathbb{R},$$

and the condition in Equation (5.2.2) and Equation (5.2.3), then

$$\sup_{t>1} \frac{1}{t} \int_0^t E\Big(\frac{|\psi'(X_u)|}{\psi(X_u)} \theta_u\Big) du < \infty.$$

$$\begin{aligned}
d(1+\theta_t) &= d\theta_t = \theta_t\Big(q - \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}\theta_t\Big)dt \\
&= q\theta_t dt - \frac{1}{\sqrt{2\pi}}\frac{(1+\theta_t)|\psi(X_t)|\theta_t}{\psi(X_t)}dt + \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}\theta_t dt \\
\Rightarrow d(1+\theta_t) + \frac{1}{\sqrt{2\pi}}\frac{(1+\theta_t)|\psi'(X_t)|\theta_t}{\psi(X_t)}dt &= q\theta_t dt + \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}\theta_t dt \\
\Rightarrow \frac{d(1+\theta_t)}{1+\theta_t} + \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}\theta_t dt &= \frac{\theta_t}{1+\theta_t}\Big(q + \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}\Big)dt \\
&\leq \Big(q + \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}\Big)dt \\
\Rightarrow \log\frac{1+\theta_t}{1+\theta_0} + \frac{1}{\sqrt{2\pi}}\int_0^t \frac{|\psi'(X_u)|}{\psi(X_u)}\theta_u du &\leq qt + \frac{1}{\sqrt{2\pi}}\int_0^t \frac{|\psi'(X_u)|}{\psi(X_u)}du \\
\Rightarrow \frac{1}{t}\int_0^t \frac{|\psi'(X_u)|}{\psi(X_u)}\theta_u du &\leq \sqrt{2\pi}q + \frac{1}{t}\int_0^t \frac{|\psi'(X_u)|}{\psi(X_u)}du + \sqrt{2\pi}\frac{\log(1+\theta_0)}{t} \\
&\leq \sqrt{2\pi}q + \frac{1}{t}\int_0^t \Big(a|X_u| + b\Big)du + \sqrt{2\pi}\frac{\log(1+\theta_0)}{t} \\
\Rightarrow \frac{1}{t}\int_0^t E\Big(\frac{|\psi'(X_u)|}{\psi(X_u)}\theta_u\Big)du &\leq \sqrt{2\pi}q + \frac{1}{t}\int_0^t E\Big(a|X_u| + b\Big)du + \sqrt{2\pi}\frac{E(\log(1+\theta_0))}{t},
\end{aligned}$$

using the growth condition in Equation (5.2.1). From Equation (5.2.25) in Theorem 7 the moments of $|X_t|$ are uniformly bounded in $t > 0$. Therefore we have that

$$\sup_{t>1}\frac{1}{t}\int_0^t E(\frac{|\psi(X_u)|}{\psi(X_u)}\theta_u)du \;<\; C.$$

■

Now we prove the main theorem of this section.

THEOREM 9. Under the hypothesis of the above lemma we have that

$$\sup_{t>1}\frac{1}{t}\int\limits_0^t E(\theta_u^{\frac{k}{2}})du \quad < \quad C, \text{ for every } k\in\mathbb{N}. \tag{5.2.34}$$

Proof: The proof is done by mathematical induction. We first prove that the hypothesis is true for $k=1$.

Recall that,

$$d\eta_t = (-q\eta_t + \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)})dt$$

where $\eta_t = \frac{1}{\theta_t}$.

Applying Itô's lemma to $Y_t = X_t^2\eta_t^{2-k/2}$, with $k\in\mathbb{N}$, we get

$$\begin{aligned}
dY_t &= 2X_t\eta_t^{2-k/2}dX_t + (2-k/2)X_t^2\eta_t^{1-k/2}d\eta_t + \frac{1}{2}2\eta_t^{2-k/2}(dX_t)^2\\
&= 2X_t\eta_t^{2-k/2}\Big(\frac{\psi'(X_t)}{2\psi(X_t)\eta_t^2}dt + \frac{1}{\eta_t}dW_t\Big) + (2-k/2)X_t^2\eta_t^{1-k/2}\Big(-q\eta_t dt + \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}dt\Big)\\
&+ \eta_t^{2-k/2}\eta_t^{-2}dt\\
&= \Big(X_t\frac{\psi'(X_t)}{\psi(X_t)}\eta_t^{-k/2} - q(2-k/2)X_t^2\eta_t^{2-k/2} + \frac{2-k/2}{\sqrt{2\pi}}X_t^2\frac{|\psi'(X_t)|}{\psi(X_t)}\eta_t^{1-k/2} + \eta_t^{-k/2}\Big)dt\\
&+ 2X_t\eta_t^{1-k/2}dW_t.
\end{aligned}\tag{5.2.35}$$

Thus, integrating both side from 0 to t and rearranging

$$\begin{aligned}
\int\limits_0^t \theta_s^{\frac{k}{2}}ds &= X_t^2\eta_t^{\frac{4-k}{2}} - X_0^2\eta_0^{\frac{4-k}{2}} - \int\limits_0^t X_s\frac{\psi'(X_s)}{\psi(X_s)}\theta_s^{\frac{k}{2}}ds\\
&+ \frac{(4-k)q}{2}\int\limits_0^t X_s^2\eta_s^{\frac{4-k}{2}}ds - \frac{4-k}{2\sqrt{2\pi}}\int\limits_0^t X_s^2\frac{|\psi'(X_s)|}{\psi(X_s)}\eta_s^{\frac{2-k}{2}}ds\\
&- 2\int\limits_0^t X_s\eta_s^{\frac{2-k}{2}}dW_s.
\end{aligned}$$

This implies

$$\begin{aligned}\frac{1}{t}\int_0^t \theta_s^{\frac{k}{2}}ds &= \frac{1}{t}(X_t^2\eta_t^{\frac{4-k}{2}} - X_0^2\eta_0^{\frac{4-k}{2}}) - \frac{1}{t}\int_0^t X_s\frac{\psi'(X_s)}{\psi(X_s)}\theta_s^{\frac{k}{2}}ds \\ &+ \frac{(4-k)q}{2t}\int_0^t X_s^2\eta_s^{\frac{4-k}{2}}ds \\ &- \frac{4-k}{2t\sqrt{2\pi}}\int_0^t X_s^2\frac{|\psi'(X_s)|}{\psi(X_s)}\eta_s^{\frac{2-k}{2}}ds - \frac{2}{t}\int_0^t X_s\eta_s^{\frac{2-k}{2}}dW_s\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{1}{t}\int_0^t E(\theta_s^{\frac{k}{2}})ds &= \frac{1}{t}(E(X_t^2\eta_t^{\frac{4-k}{2}}) - E(X_0^2\eta_0^{\frac{4-k}{2}})) - \frac{1}{t}\int_0^t E(X_s\frac{\psi'(X_s)}{\psi(X_s)}\theta_s^{\frac{k}{2}})ds \\ &+ \frac{(4-k)q}{2t}\int_0^t E(X_s^2\eta_s^{\frac{4-k}{2}})ds \\ &- \frac{4-k}{2t\sqrt{2\pi}}\int_0^t X_s^2\frac{|\psi'(X_s)|}{\psi(X_s)}\eta_s^{\frac{2-k}{2}}ds - \frac{2}{t}E\int_0^t X_s\eta_s^{\frac{2-k}{2}}dW_s. \qquad (5.2.36)\end{aligned}$$

The last expectation is zero since $\int_0^t X_s\eta_s^{\frac{2-k}{2}}dW_s$ is a square integrable martingale for $k = 1, 2$ (by the finiteness of moments of X_t and η_t). For $k \geq 3$ we have

$$\begin{aligned}E\Big(\int_0^t X_s\eta_s^{\frac{2-k}{2}}dW_s\Big)^2 = E\Big(\int_0^t X_s\theta_s^{\frac{k-2}{2}}dW_s\Big)^2 &= \int_0^t E(X_s^2\theta_s^{k-2})ds \leq \int_0^t \sqrt{E(X_s^4)E\theta_s^{2(k-2)}}ds \\ &\leq \sqrt{\int_0^t E(X_s^4)ds}\sqrt{\int_0^t E(\theta_s^{2(k-2)})ds},\end{aligned}$$

where the last two inequalities follow from Holder by taking $p = q = \frac{1}{2}$. The first term $E(X_s^4)$ is finite for any $t > 0$. For the second term note that from the SDE of θ

$$\begin{aligned} d\theta_t &= \theta_t\Big(q - \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_t)|}{\psi(X_t)}\theta_t\Big)dt \leq q\theta_t dt \\ \Rightarrow \theta_t^{2(k-2)} &\leq \theta_0^{2(k-2)}e^{2qt(k-2)}, \end{aligned} \tag{5.2.37}$$

for any $k \in \mathbb{N}$. Therefore $\int_0^t E(\theta_s^{2(k-2)})ds < \infty$ for any $t \in [0, T], T > 0$. Therefore for any $t > 0$, $\int_0^t X_s \theta_s^{\frac{k-2}{2}} dW_s$ is a square integrable martingale on $[0, T]$and hence its expectation is zero. Therefore removing that term we have for any $k \in \mathbb{N}$,

$$\begin{aligned} \frac{1}{t}\int_0^t E(\theta_s^{\frac{k}{2}})ds &= \frac{1}{t}(E(X_t^2\eta_t^{\frac{4-k}{2}}) - E(X_0^2\eta_0^{\frac{4-k}{2}})) - \frac{1}{t}\int_0^t E(X_s\frac{\psi'(X_s)}{\psi(X_s)}\theta_s^{\frac{k}{2}})ds \\ &+ \frac{(4-k)q}{2t}\int_0^t E(X_s^2\eta_s^{\frac{4-k}{2}})ds \\ &- \underbrace{\frac{4-k}{2t\sqrt{2\pi}}\int_0^t E\Big(X_s^2\frac{|\psi'(X_s)|}{\psi(X_s)}\eta_s^{\frac{2-k}{2}}\Big)ds}_{A_1}\,. \end{aligned} \tag{5.2.38}$$

Now since $A_1 := \frac{4-k}{2t\sqrt{2\pi}}X_s^2\frac{|\psi'(X_s)|}{\psi(X_s)}\eta_s^{\frac{2-k}{2}} \geq 0$, almost surely for $1 \leq k \leq 4$ we have for these values of k

$$\begin{aligned} \sup_{t>1}\frac{1}{t}\int_0^t E(\theta_s^{\frac{k}{2}})ds &\leq \underbrace{\sup_{t>1}\frac{1}{t}(E(X_t^2\eta_t^{\frac{4-k}{2}}) - E(X_0^2\eta_0^{\frac{4-k}{2}}))}_{B_1} + \underbrace{\sup_{t>1}\frac{1}{t}\int_0^t E(|X_s\frac{\psi'(X_s)}{\psi(X_s)}\theta_s^{\frac{k}{2}}|)ds}_{C_1} \\ &+ \underbrace{\frac{(4-k)q}{2}\sup_{t>1}\frac{1}{t}\int_0^t E(X_s^2\eta_s^{\frac{4-k}{2}})ds}_{D_1}\,. \end{aligned} \tag{5.2.39}$$

For the term

$$C_1 := \sup_{t>1}\frac{1}{t}\int_0^t E(|X_s|Z_s\theta_s^{\frac{k}{2}}ds),$$

where $Z_s = \frac{|\psi'(X_s)|}{\psi(X_s)}$, we have for any $k \in \mathbb{N}$,

$$\begin{aligned}
\frac{1}{t}\int_0^t |X_s| Z_s \theta_s^{\frac{k}{2}} ds &= \frac{1}{t}\int_0^t |Z_s|^{\frac{k}{k+1}} \theta_s^{\frac{k}{2}} |X_s| Z_s^{\frac{1}{k+1}} ds \\
&\leq \Big(\frac{1}{t}\int_0^t Z_s \theta_s^{\frac{k+1}{2}} ds\Big)^{\frac{k}{k+1}} \Big(\frac{1}{t}\int_0^t |X_s|^{k+1} Z_s ds\Big)^{\frac{1}{k+1}}, \qquad (5.2.40)
\end{aligned}$$

which follows from the Holder's inequality with $p = \frac{k+1}{k}$ and $q = k+1$. Therefore,

$$\begin{aligned}
E\Big(\frac{1}{t}\int_0^t |X_s| Z_s \theta_s^{\frac{k}{2}}\Big) &\leq E\Big(\Big(\frac{1}{t}\int_0^t Z_s \theta_s^{\frac{k+1}{2}} ds\Big)^{\frac{k}{k+1}} \Big(\frac{1}{t}\int_0^t |X_s|^{k+1} Z_s ds\Big)^{\frac{1}{k+1}}\Big) \\
&\leq \Big(E\Big(\frac{1}{t}\int_0^t Z_s \theta_s^{\frac{k+1}{2}} ds\Big)\Big)^{\frac{k}{k+1}} \times \Big(E\Big(\frac{1}{t}\int_0^t |X_s|^{k+1} Z_s ds\Big)\Big)^{\frac{1}{k+1}} \\
&= \Big(\frac{1}{t}\int_0^t E(Z_s \theta_s^{\frac{k+1}{2}}) ds\Big)^{\frac{k}{k+1}} \Big(\frac{1}{t}\int_0^t E(|X_s|^{k+1}) Z_s ds\Big)^{\frac{1}{k+1}},
\end{aligned}$$

where the last inequality follows from Holder's inequality with $p = \frac{k+1}{k}$ and $q = k+1$. Therefore

$$\begin{aligned}
C_1 = \sup_{t>1} E\Big(\frac{1}{t}\int_0^t |X_s| Z_s \theta_s^{\frac{k}{2}}\Big) &\leq \Big(\sup_{t>1} \frac{1}{t}\int_0^t E(Z_s \theta_s^{\frac{k+1}{2}}) ds\Big)^{\frac{k}{k+1}} \\
&\quad \Big(\sup_{t>1} \frac{1}{t}\int_0^t E(|X_s|^{k+1}) Z_s ds\Big)^{\frac{1}{k+1}}, \qquad (5.2.41)
\end{aligned}$$

From the hypothesis of the Theorem we have

$$\begin{aligned}
\frac{1}{t}\int_0^t E(|X_s|^{k+1} Z_s) ds &\leq \frac{1}{t}\int_0^t E(|X_s|^{k+1}(a|X_s| + b)) ds, \\
&= \frac{1}{t}\Big(\int_0^t \Big(aE|X_s|^{k+2} + bE|X_s|^{k+1}\Big) ds\Big). \qquad (5.2.42)
\end{aligned}$$

This implies from Equation (5.2.41) that

$$C_1 \leq \underbrace{\sup_{t>1}\Big(\frac{1}{t}\int_0^t E(Z_s\theta_s^{\frac{k+1}{2}})ds\Big)^{\frac{k}{k+1}}}_{E_1} \times \sup_{t>1}\Big(\frac{1}{t}\int_0^t (aE|X_s|^{k+2} + bE|X_s|^{k+1})ds\Big)^{\frac{k}{k+1}} \tag{5.2.43}$$

Note that for $k = 1$, Lemma 30 implies that E_1 is finite. This with the fact that the moments of X_t are uniformly bounded (by Equation (5.2.25) of Theorem 7) implies that C_1 is finite for $k = 1$. Again the terms B_1 and D_1 in Equation (5.2.39) is finite by the finiteness of moments of X_t and η_t and by a simple application of the Cauchy-Schwartz inequality. Hence the Theorem is true for $k = 1$. This completes the first step of the induction.

Assume that the hypothesis is true for $m = k - 1$, that is,

$$\sup_{t>1}\frac{1}{t}\int_0^t E(\theta_s^{\frac{k-1}{2}})ds \leq C < \infty. \tag{5.2.44}$$

We proceed to show that the hypothesis is true for $m = k$, that is

$$\sup_{t>1}\frac{1}{t}\int_0^t E(\theta_s^{\frac{k}{2}})ds \leq C < \infty. \tag{5.2.45}$$

For the term E_1 in Equation (5.2.43) we have that for any $k > 1$,

$$\begin{aligned} d\theta_t^{\frac{k-1}{2}} = \frac{k-1}{2}\theta_t^{\frac{k-1}{2}-1}d\theta_t &= \frac{k-1}{2}\theta_t^{\frac{k-1}{2}}\Big(q - \frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_s)|}{\psi(X_s)}\theta_t\Big)dt \\ &= \frac{q(k-1)}{2}\theta_t^{\frac{k-1}{2}}dt - \frac{k-1}{2}\frac{1}{\sqrt{2\pi}}\frac{|\psi'(X_s)|}{\psi(X_s)}\theta_t^{\frac{k+1}{2}}dt \end{aligned} \tag{5.2.46}$$

$$\begin{aligned}
\Rightarrow \frac{k-1}{2\sqrt{2\pi}}\int_0^t Z_s\theta_s^{\frac{k+1}{2}}ds &= \int_0^t \frac{q(k-1)}{2}\theta_s^{\frac{k-1}{2}}ds - (\theta_t^{\frac{k-1}{2}} - \theta_0^{\frac{k-1}{2}}) \ \ (\text{recall } Z_s = \tfrac{|\psi'(X_s)|}{\psi(X_s)}) \\
\Rightarrow \frac{2\sqrt{2\pi}}{k-1}E(\theta_t^{\frac{k-1}{2}}) + \int_0^t E(Z_s\theta_s^{\frac{k+1}{2}})ds &= q\sqrt{2\pi}\int_0^t E(\theta_s^{\frac{k-1}{2}})ds \\
&+ \frac{2\sqrt{2\pi}}{k-1}E(\theta_0^{\frac{k-1}{2}}) \\
\Rightarrow E_1 = \sup_{t>1}\frac{1}{t}\int_0^t E\Big(Z_s\theta_s^{\frac{k+1}{2}}ds\Big) &\le q\sqrt{2\pi}\sup_{t>1}\frac{1}{t}\int_0^t E\Big(\theta_s^{\frac{k-1}{2}}ds\Big) \\
&+ \frac{2\sqrt{2\pi}}{k-1}\sup_{t>1}\frac{1}{t}E\Big(\theta_0^{\frac{k-1}{2}}\Big). \qquad (5.2.47)
\end{aligned}$$

Plugging (5.2.47) and (5.2.42) in (5.2.41) we get

$$\begin{aligned}
C_1 = \sup_{t>1}\frac{1}{t}\int_0^t E(|X_s|Z_s\theta_s^{\frac{k}{2}})ds &\le \sup_{t>1}\Big(q\sqrt{2\pi}\sup_{t>1}\frac{1}{t}\int_0^t E(\theta_s^{\frac{k-1}{2}})ds \\
&+ \frac{2\sqrt{2\pi}}{k-1}\frac{1}{t}E(\theta_0^{\frac{k-1}{2}})\Big)^{\frac{k}{k+1}} \\
&\times \Big(\sup_{t>1}\frac{1}{t}\int_0^t (aE|X_s|^{k+2} + bE|X_s|^{k+1})ds\Big)^{\frac{1}{k+1}}.
\end{aligned}$$

(5.2.48)

Therefore C_1 is finite by assumption (5.2.44) and finiteness of moments of X_t.
And finally plugging (5.2.48) in (5.2.39) we get for $2 \leq k \leq 4$

$$
\begin{aligned}
\sup_{t>1} \frac{1}{t} \int_0^t E(\theta_s^{\frac{k}{2}}) ds &\leq \sup_{t>1} \frac{1}{t} E(X_t^2 \eta_t^{\frac{4-k}{2}} - X_0^2 \eta_0^{\frac{4-k}{2}}) \\
&+ \Bigg(q\sqrt{2\pi} \sup_{t>1} \frac{1}{t} \int_0^t E(\theta_s^{\frac{k-1}{2}}) ds \\
&+ \frac{2}{k-1} \sqrt{2\pi} \sup_{t>1} \frac{1}{t} E(\theta_0^{\frac{k-1}{2}}) \Bigg)^{\frac{k}{k+1}} \\
&\times \left(\sup_{t>1} \frac{1}{t} \int_0^t (aE|X_s|^{k+2} + bE|X_s|^{k+1}) ds \right)^{\frac{1}{k+1}} \\
&+ \frac{(4-k)q}{2} \sup_{t>1} \frac{1}{t} \int_0^t E(X_s^2 \eta_s^{\frac{4-k}{2}}) ds. \qquad (5.2.49)
\end{aligned}
$$

By the assumption (5.2.44) and the uniform boundedness of moments of X_t and η_t all terms in the R.H.S of the above equation is finite. However for $k > 4$ we have the negative values of A_1 in Equation (5.2.39) and the term D_1 is negative. Therefore

$$
\begin{aligned}
\sup_{t>1} \frac{1}{t} \int_0^t E(\theta_s^{\frac{k}{2}}) ds &\leq \underbrace{\sup_{t>1} \frac{1}{t} E(X_t^2 \theta_t^{\frac{k-4}{2}}}_{G_1} - X_0^2 \theta_0^{\frac{k-4}{2}}) \\
&+ \frac{k-1}{2} \Bigg(q\sqrt{2\pi} \sup_{t>1} \frac{1}{t} \int_0^t E(\theta_s^{\frac{k-1}{2}}) ds \\
&+ \sqrt{2\pi} \sup_{t>1} \frac{1}{t} E(\theta_0^{\frac{k-1}{2}}) \Bigg)^{\frac{k}{k+1}} \\
&\times \left(\sup_{t>1} \frac{1}{t} \int_0^t (aE|X_s|^{k+2} + bE|X_s|^{k+1}) ds \right)^{\frac{1}{k+1}} \\
&+ \underbrace{\frac{k-4}{2\sqrt{2\pi}} \sup_{t>1} \frac{1}{t} \int_0^t E(X_s^2 \frac{|\psi'(X_s)|}{\psi(X_s)} \theta_s^{\frac{k-2}{2}} ds)}_{F_1}. \qquad (5.2.50)
\end{aligned}
$$

Now for the term

$$F_1 := \frac{k-4}{2\sqrt{2\pi}} \sup_{t>1} \frac{1}{t} \int_0^t E(X_s^2 \frac{|\psi'(X_s)|}{\psi(X_s)} \theta_s^{\frac{k-2}{2}} ds)$$

note that

$$\begin{aligned} \int_0^t X_s^2 \frac{|\psi'(X_s)|}{\psi(X_s)} \theta_s^{\frac{k-2}{2}} ds &= \int_0^t X_s^2 Z_s \theta_s^{\frac{k-2}{2}} ds \\ &= \int_0^t \underbrace{Z_s^{\frac{k-2}{k+1}} \theta_s^{\frac{k-2}{2}}} \underbrace{|X_s|^2 Z_s^{\frac{3}{k+1}}} ds \\ &\le \Big(\int_0^t Z_s \theta_s^{\frac{k+1}{2}} ds\Big)^{\frac{k-2}{k+1}} \Big(\int_0^t |X_s|^{\frac{2(k+1)}{3}} Z_s ds\Big)^{\frac{3}{k+1}}, \end{aligned}$$

which follows from Holder's inequality by taking $p = \frac{k+1}{k-2}$ and $q = \frac{k+1}{3}$. This implies that

$$\begin{aligned} \sup_{t>1} \frac{1}{t} E\Big(\int_0^t X_s^2 Z_s \theta_s^{\frac{k-2}{2}} ds\Big) &\le \sup_{t>1} \Big(\frac{1}{t} \int_0^t E(Z_s \theta_s^{\frac{k+1}{2}}) ds\Big)^{\frac{k-2}{k+1}} \\ &\times \sup_{t>1} \Big(\frac{1}{t} \int_0^t E(|X_s|^{\frac{2(k+1)}{3}} Z_s) ds\Big)^{\frac{k-2}{k+1}}, \end{aligned}$$

which follows by a similar application of Holder's inequality with the same value of p and q. Now it has been shown in Equation (5.2.47) that by induction hypothesis,

$$\sup_{t>1} \Big(\frac{1}{t} \int_0^t E(Z_s \theta_s^{\frac{k+1}{2}}) ds\Big) \le C < \infty.$$

Applying condition in Equation (5.2.1) we have that

$$\begin{aligned} E(|X_s|^{\frac{2(k+1)}{3}} Z_s) &\le E\Big(|X_s|^{\frac{2(k+1)}{3}} (a|X_s| + b)\Big) \\ &= E\Big(|X_s|^{\frac{2k+5}{3}} + b|X_s|^{\frac{2(k+1)}{3}}\Big) \\ \Rightarrow \sup_{t>1} \frac{1}{t} \int_0^t E(|X_s|^{\frac{2(k+1)}{3}} Z_s) ds &< \infty, \end{aligned} \tag{5.2.51}$$

which follows form the uniform boundedness of all moments of X_t. For the term $G_1 = \sup\limits_{t>1} \frac{1}{t} E(X_t^2 \theta_t^{\frac{k-4}{2}})$, an application of Young's inequality with $p = \frac{k-1}{k-4}$ and $q = \frac{k-1}{3}$ gives

$$\begin{aligned} X_t^2 \theta_t^{\frac{k-4}{2}} &\leq \frac{3}{k-1} X_t^{\frac{2(k-1)}{3}} + \frac{k-4}{k-1} \theta_t^{\frac{k-1}{2}} \\ \Rightarrow \sup_{t>1} \frac{1}{t} E(X_t^2 \theta_t^{\frac{k-4}{2}}) &\leq \frac{3}{k-1} \sup_{t>1} \frac{1}{t} E(X_t^{\frac{2(k-1)}{3}}) + \frac{k-4}{k-1} \sup_{t>1} \frac{1}{t} E(\theta_t^{\frac{k-1}{2}}). \end{aligned} \tag{5.2.52}$$

Now the first quantity on the right hand side of the above inequality is finite by the finiteness of the moments of X_t. For the second term we have by (5.2.46)

$$\begin{aligned} d\theta_t^{\frac{k-1}{2}} &= \frac{q(k-1)}{2} \theta_t^{\frac{k-1}{2}} dt - \frac{k-1}{2} \frac{1}{\sqrt{2\pi}} \frac{|\psi'(X_t)|}{\psi(X_t)} \theta_t^{\frac{k+1}{2}} dt \\ &\leq \frac{q(k-1)}{2} \theta_t^{\frac{k-1}{2}} dt \\ \Rightarrow \theta_t^{\frac{k-1}{2}} &\leq \theta_0^{\frac{k-1}{2}} + \frac{q(k-1)}{2} \int_0^t \theta_s^{\frac{k-1}{2}} ds \\ \Rightarrow \sup_{t>1} \frac{1}{t} E(\theta_t^{\frac{k-1}{2}}) &\leq \sup_{t>1} \frac{1}{t} E(\theta_0^{\frac{k-1}{2}}) + \frac{q(k-1)}{2} \sup_{t>1} \frac{1}{t} \int_0^t E(\theta_s^{\frac{k-1}{2}}) ds < \infty, \end{aligned}$$

by the induction hypothesis (5.2.44) and the hypothesis of the theorem. This proves that

$$\sup_{t>1} \frac{1}{t} E(X_t^2 \theta_t^{\frac{k-4}{2}}) < \infty$$

which in turn proves that all terms in the R.H.S of Equation (5.2.50) is finite. This proves, for any $k > 4$ Equation (5.2.45) holds. This proves the theorem. ■.

Remark 31. To find the almost sure bound to the growth of $\frac{1}{t} \int_0^t \theta_s ds$ (instead of the expected value) we need to find the almost sure growth of X_t, $t > 0$ or $\frac{1}{t} \int_0^t X_s ds$. However, this is available when the two dimensional diffusion has an unique invariant distribution, say $\mu(\cdot)$ and it is jointly ergodic i.e.,

$$\frac{1}{t} \int_0^t f(X_s, \theta_s) ds \to \int f d\mu,$$

whenever $\int |f| d\mu < \infty$.

REMARK 32. For the AMCMC with target density $\psi(\cdot)$ we have

$$\begin{aligned}
\theta_t &= \frac{e^{qt}}{\eta_0 + \frac{1}{\sqrt{2\pi}} \int\limits_0^t e^{qs} \frac{|\psi'(X_s)|}{\psi(X_s)}} ds \\
\Rightarrow \theta_t^2 &\geq \frac{e^{2qt}}{2\eta_0^2 + \frac{e^{qt}-1}{\pi q^2} \int\limits_0^t e^{qs} E(\frac{\psi'(X_s)}{\psi(X_s)})^2 ds}
\end{aligned}$$

Since $\frac{|\psi'(X_s)|}{\psi(X_s)} < a|X_s| + b$, we have

$$\begin{aligned}
E(\theta_t^2) &\geq \frac{e^{2qt}}{2\eta_0^2 + \frac{e^{qt}-1}{\pi q} \int\limits_0^t e^{qs} E(a|X_s| + b)^2 ds}, \\
&\geq \frac{e^{2qt}}{2\eta_0^2 + \frac{e^{qt}-1}{\pi q} \int\limits_0^t e^{qs} \Big(a^2 M_2 + b^2 + 2abM_1\Big) ds},
\end{aligned}$$

where M_2 and M_1 are uniform bounds for $\sup\limits_{t>0} E(X_t^2)$ and $\sup\limits_{t>0} E(|X_t|)$. Therefore

$$\begin{aligned}
\liminf_{t\to\infty} E(\theta_t^2) &\geq \frac{\pi q^2}{b^2 + a^2 M_2 + 2abM_1} \\
\liminf_{t\to\infty} \frac{1}{t} \int\limits_0^t E(\theta_s^2) ds &\geq \frac{\pi q^2}{b^2 + a^2 M_2 + 2abM_1} \quad .
\end{aligned}$$

This gives a lower bound to the growth of θ_t.

REMARK 33. The interval $(-\infty, \infty)$ is a natural boundary for the SDE corresponding to the SMCMC, i.e, Equation (3.4.2). For a one dimensional diffusion whose generator is given by

$$b(x)\frac{\partial}{\partial x} + a(x)\frac{\partial^2}{\partial x^2},$$

an interval (r_i, r_j) is natural if it is not regular, exit or an entrance boundary, see the definitions in pg.516 of [24]. Indeed, define $W(x) = exp(-\int_{x_0}^x b(s)a^{-1}(s)ds)$, for some $x_0 \in (-\infty, \infty)$. For the general target distribution $\psi(\cdot)$, $W(x) = exp(-\int\limits_{x_0}^{x} \frac{\psi'(s)}{\psi(s)} ds)$. In the standard Normal case,

$$W(x) = exp(\int_{x_0}^{x} sds) = exp(\frac{1}{2}(x^2 - x_0^2))$$

which is not integrable in (x_0, ∞). Again,

$$\begin{aligned} W(x)\int_{x_0}^{x} a^{-1}(s)W^{-1}(s)ds &= \frac{1}{\theta_0}exp(\frac{x^2}{2})\int_{x_0}^{x} e^{-\frac{s^2}{2}}ds \\ &= \frac{\sqrt{2\pi}}{\theta_0}exp(\frac{x^2}{2})(\Phi(x) - \Phi(x_0)). \end{aligned}$$

The quantity on the RHS of the last equation is not integrable in $\mathbb{R}$. Indeed, since for large $x, exp(\frac{x^2}{2})\Phi(x) \approx exp(\frac{x^2}{2})\frac{\phi(x)}{x} = \frac{1}{x}$ it is not integrable in $(-\infty, \infty)$.

5.2.4 Heavy tailed target distribution

In this section we consider densities satisfying condition in (5.2.4). We showed that in such cases the m.g.f does not exists. We give some examples.

Example 1: (Standard Cauchy density) For the standard Cauchy density $\psi(x) = \frac{1}{\pi}\frac{1}{1+x^2}$. Hence the LHS of (5.2.2) is

$$\begin{aligned} x\frac{\psi'(x)}{\psi(x)} &= -\frac{2x^2}{1+x^2} \\ &= -2\Big(1 - \frac{1}{1+x^2}\Big) \to -2 \text{ as } |x| \to \infty. \end{aligned}$$

■

Example 2: (Pareto tails) Consider the Pareto distribution with tails following a power law i.e., $\overline{F}(x) := 1 - F(x) \sim \frac{1}{x^\alpha}$ where $\alpha > 0$. This would imply that $\psi(x) \sim x^{-(1+\alpha)}$. Then the LHS of (5.2.2) takes the value

$$x\frac{\psi'(x)}{\psi(x)} \sim -(1+\alpha),$$

which tends to $-(1+\alpha)$ as $|x| \to \infty$. So the Pareto distribution satisfy condition (5.2.4). ■

If the target density is Standard Cauchy then $\psi(x) = \frac{1}{\pi(1+x^2)} \Rightarrow \frac{\psi'(x)}{\psi(x)} = -\frac{2x}{1+x^2}$. Therefore the SDE of (X_t, θ_t) will be

$$\begin{aligned} dX_t &= -\theta_t^2 \frac{X_t}{1+X_t^2} dt + \theta_t dW_t, & (5.2.53)\\ d\theta_t &= \theta_t \Big(q - \sqrt{\frac{2}{\pi}} \frac{|X_t|}{1+X_t^2} dt \Big) & (5.2.54) \end{aligned}$$

Define $V_t = 1 + X_t^2$ and $C_t = \int\limits_0^t \theta_s^2 V_s^{-1} ds$. Then we have the following lemma.

LEMMA 31. Let X_t and θ_t be the solutions of the SDE (5.2.53) and (5.2.54). If X_0 and θ_0 are independent and the m.g.f of θ_0^2 exists for all $t \in \mathbb{R}$ then

$$E\Big(e^{-C_t} \int\limits_0^t e^{C_s} X_s \theta_s dW_s\Big) = 0,$$

where $V_t = 1 + X_t^2$ and $C_t = \int\limits_0^t \theta_s^2 V_s^{-1} ds$.

Proof: Define $Z_t = e^{-C_t} Y_t$ where $Y_t = \int\limits_0^t e^{C_s} X_s \theta_s dW_s$. Then it can be proved using methods from Lemma 12 that Z_t is symmetric about zero. To claim that its expectation is zero we need to prove that $E(|Z_t|) < \infty$, $\forall t$. Since $E|Z_t| < \sqrt{E(e^{-2c_t})E(Y_t^2)}$, thus, for each $T > 0$, it is sufficient to show that

$$E(\int\limits_0^t e^{2C_s} \theta_s^2 X_s^2 ds) < \infty \text{ for all } t \in [0, T]. \tag{5.2.55}$$

From the SDE of θ_t (5.2.54) we have

$$\begin{aligned} d\theta_t^2 = 2\theta_t d\theta_t &= 2\theta_t^2 \Big(q - \sqrt{\frac{2}{\pi}} \frac{|X_t|}{1+X_t^2} dt \Big) \le 2q\theta_t^2 \\ \Rightarrow \theta_t^2 &\le \theta_0^2 e^{2qt} & (5.2.56) \end{aligned}$$

From the SDE of X_t (5.2.53) we have

$$\begin{aligned} X_t &= X_0 - \int_0^t \theta_s^2 \frac{X_s}{1+X_s^2} + \int_0^t \theta_s dW_s \\ \Rightarrow X_t^2 &\le D_2\Big(X_0^2 + (\int_0^t \frac{\theta_s^2}{2}\frac{2X_s}{1+X_s^2}ds)^2 + (\int_0^t \theta_s dW_s)^2\Big) \\ &\le D_2\Big(X_0^2 + \frac{1}{4}(\int_0^t \theta_0^2 e^{2qs}ds)^2 + (\int_0^t \theta_s dW_s)^2\Big) \\ &= D_2\Big(X_0^2 + \frac{\theta_0^4}{4}(\frac{e^{2qt}-1}{2q})^2 + (\int_0^t \theta_s dW_s)^2\Big), \end{aligned} \tag{5.2.57}$$

for some $D_2 > 0$. Also

$$e^{2C_s} \le e^{2\int_0^s \theta_u^2 du} \le e^{\theta_0^2 e^{(2qs-1)/q}}.$$

Therefore

$$\begin{aligned} E(\int_0^t e^{2C_s}\theta_s^2 X_s^2 ds) &\le E\Big(\int_0^t e^{\theta_0^2 e^{(2qs-1)/q}} \theta_0^2 e^{2qs} X_s^2 ds\Big) \\ &\le D_2 E\Big(\theta_0^2 \int_0^t e^{\theta_0^2 e^{(2qs-1)/q}+2qs}\Big(X_0^2 + \frac{\theta_0^4}{4}(\frac{e^{2qs}-1}{2q})^2 \\ &\quad +(\int_0^s \theta_u dW_u)^2 ds\Big)\Big) \\ &\le D_2\Big(E\Big(\theta_0^2 X_0^2 \int_0^t e^{\theta_0^2 e^{(2qs-1)/q}+2qs}ds\Big) + \frac{1}{4}E\Big(\theta_0^6 \int_0^t (\frac{e^{2qs}-1}{2q})^2 e^{\theta_0^2 e^{(2qs-1)/q}+2qs}ds\Big) \\ &+ \int_0^t E\Big(e^{\theta_0^2 e^{(2qs-1)/q}+2qs}\Big(\theta_0 \int_0^s \theta_u dW_u\Big)^2 ds\Big)\Big) \end{aligned} \tag{5.2.58}$$

The expectation in the last integral of (5.2.58) can be written as

$$E\Big(\theta_0^2 e^{\theta_0^2 e^{\frac{2qs-1}{2}+2qs}}\big(\int_0^s \theta_u dW_u\big)^2\Big) \;\leq\; \sqrt{E\Big(\theta_0^4 e^{2\theta_0^2 e^{\frac{2qs-1}{2}+2qs}}\Big)E\Big(\int_0^s \theta_u dW_u\Big)^4}.$$

Since θ_0 is a positive random variable and its m.g.f exists for all $t \in \mathbb{R}$ the m.g.f of any power of θ_0 also exist for any $t \in \mathbb{R}$. Therefore by the Cauchy Schwarz inequality the first expectation in the RHS of the last inequality is finite. For the second expectation note that since $M_s := \int_0^s \theta_u dW_u$ is a square integrable martingale we apply the BDG inequality to get

$$\begin{aligned} E\big(\int_0^s \theta_u dW_u\big)^4 &\leq D_4 E\big(\int_0^s \theta_u^2 du\big)^2, \text{ for some } D_4 < \infty, \\ &\leq D_4 E\big(\int_0^s \theta_0^2 e^{2qs} ds\big) = D_4 \frac{e^{2qs}-1}{2q} E(\theta_0^2), \end{aligned} \tag{5.2.59}$$

where $D_4 \in (0, \infty)$ is a constant. Hence plugging the value obtained in (5.2.59) to (5.2.58) and applying the hypothesis in the Lemma to (5.2.58) we see that $E\int_0^t e^{2Cs}\theta_s^2 X_s^2 ds < \infty$, for any $t \in [0, T], T < \infty$. This proves the lemma. ■

The following theorem gives a sufficient condition for the limiting second moment to be finite.

THEOREM 10. If $\psi(\cdot)$ is the standard Cauchy density then $\limsup_{t\to\infty} E(X_t^2) < \infty$ if the hypothesis of Lemma 31 holds.

Proof: Applying Itô's lemma to $V_t := 1 + X_t^2$ we have

$$
\begin{aligned}
dV_t &= 2X_t dX_t + 2\frac{1}{2}(dX_t)^2 \\
&= 2X_t\Big(\frac{\theta_t^2}{2}\frac{\psi'(X_t)}{\psi(X_t)})dt + \theta_t dW_t\Big) + \theta_t^2 dt \\
&= 2X_t\Big(-\theta_t^2 \frac{X_t}{1+X_t^2}dt + \theta_t dW_t\Big) + \theta_t^2 dt \\
&= -\frac{2X_t^2\theta_t^2}{1+X_t^2}dt + \theta_t^2 dt + 2X_t\theta_t dW_t \\
&= -2\theta_t^2(\frac{1+X_t^2-1}{1+X_t^2})dt + \theta_t^2 dt + 2X_t\theta_t dW_t \\
&= -2\theta_t^2(1-\frac{1}{1+X_t^2})dt + \theta_t^2 dt + 2X_t\theta_t dW_t \\
\Rightarrow dV_t &= -2\theta_t^2 dt + \frac{2\theta_t^2}{1+X_t^2}dt + \theta_t^2 dt + 2X_t\theta_t dW_t \\
\Rightarrow dV_t + \theta_t^2 dt &= 2\theta_t^2 V_t^{-1}dt + 2X_t\theta_t dW_t
\end{aligned}
$$

Multiplying by the integrating factor $e^{\int_0^t \theta_s^2 V_s^{-1} ds} := e^{C_t}$ on both sides we get

$$
d\Big(V_t e^{C_t}\Big) = e^{C_t}\Big(2\theta_t^2 V_t^{-1}dt + 2X_t\theta_t dW_t\Big)
$$

Integrating from 0 to t on both sides

$$
\begin{aligned}
V_t &= V_0 e^{-C_t} + 2e^{-C_t}\int_0^t e^{C_s}\theta_s^2 V_s^{-1} ds + 2e^{-C_t}\int_0^t e^{C_s} X_s\theta_s dW_s \\
&= V_0 e^{-C_t} + 2e^{-C_t}(e^{C_t}-1) + 2e^{-C_t}\int_0^t e^{C_s} X_s\theta_s dW_s \text{ (since } d(e^{C_t}) = e^{C_t}\theta_t^2 V_t^{-1}dt) \\
&= V_0 e^{-C_t} + 2(1-e^{-C_t}) + 2e^{-C_t}\int_0^t e^{C_s} X_s\theta_s dW_s \\
\Rightarrow E(V_t) &= E(V_0 e^{-C_t}) + 2E(1-e^{-C_t}) + 2E\Big(e^{-C_t}\int_0^t e^{C_s} X_s\theta_s dW_s\Big). \qquad (5.2.60)
\end{aligned}
$$

The third expectation is zero by Lemma 31. Therefore from (5.2.60)

$$\begin{aligned} E(V_t) &\leq K_1 < \infty \text{ for all } t > 0, \text{and hence,} \\ \limsup_{t\to\infty} E(X_t^2) &< K \text{ for some } K < \infty. \end{aligned}$$

This proves the theorem. ∎

LEMMA 32. Under the hypothesis of Lemma 31 if $E(\eta_0^{2k}) < \infty$ then

$$\sup_{t>0} E(\eta_t^{2k}) < \infty.$$

Proof: Following the proof for the Normal distribution in Lemma 15 we can conclude that the moments of $\eta_t := \frac{1}{\theta_t}$ are finite. Indeed,

$$\begin{aligned} d\theta_t &= \theta_t\Big(q - \theta_t\sqrt{\frac{2}{\pi}}\frac{|X_t|}{1+X_t^2}\Big)dt \\ \Rightarrow d\eta_t &= -\frac{1}{\theta_t^2}d\theta_t = \Big(-q\eta_t + \sqrt{\frac{2}{\pi}}\frac{|X_t|}{1+X_t^2}\Big)dt \end{aligned}$$

Multiplying by the integrating factor e^{qt} on both sides of above we have

$$\begin{aligned} d\Big(e^{qt}\eta_t\Big) &= \sqrt{\frac{2}{\pi}}e^{qt}\frac{|X_t|}{1+X_t^2}dt \\ \Rightarrow \eta_t &= \eta_0 e^{-qt} + \sqrt{\frac{2}{\pi}}\int_0^t e^{-q(t-s)}\frac{|X_s|}{1+X_s^2}ds \\ &\leq \eta_0 e^{-qt} + \sqrt{\frac{1}{2\pi}}\int_0^t e^{-q(t-s)}ds, \text{ since } \tfrac{2|x|}{1+x^2} \leq 1, \\ \Rightarrow \eta_t^{2k} &\leq 2^{2k-1}\Big(\eta_0^{2k}e^{-2kqt} + \Big(\frac{1}{2\pi}\Big)^k\Big(\int_0^t e^{-q(t-s)}ds\Big)^{2k}\Big) \\ &\leq 2^{2k-1}\Big(\eta_0^{2k}e^{-2kqt} + \frac{1}{(2\pi)^k}\Big(\frac{1-e^{-qt}}{q}\Big)^{2k}\Big) \\ \Rightarrow \sup_{t>0} E(\eta_t^{2k}) &< M_k < \infty. \end{aligned} \tag{5.2.61}$$

This completes the proof of the lemma. ∎

Remark 34. Similar to the arguments given in Remark 28 one can conclude that if the m.g.f of X_0 and η_0 exist for $t \in \mathbb{R}$ and under the hypothesis of Lemma 31 the m.g.f. of $\sup_{t>0} E(e^{s\eta_t}) < \infty$. for all $s \in \mathbb{R}$.

Remark 35. From Theorem 10, Lemma 32 and applying the argument given in Lemma 16 of Chapter 4 we get that the process $\{(X_t, \eta_t)\}_{t>0}$ is tight. Hence, if the limiting distribution exists for this one, then its X-marginal cannot be standard Cauchy since the latter do not have moments of order one or higher. Following Stramer and Tweedie [63], SMCMC will have the same diffusion equation as in (3.4.2) and thus will have the same invariant distribution as in [63]. Since the proof of the bounds of the second moments of the diffusion $\{X_t\}$ corresponding to SMCMC is exactly similar to that of Theorem 10 (proved for AMCMC), we can conclude that the limiting distribution in this case cannot be standard Cauchy either. This means that if we simulate from the diffusions generated by the SMCMC or the AMCMC algorithms, with a fixed starting point and using the standard Cauchy as the target density, we end up in a different distribution which is not standard Cauchy. Perhaps it is due to the assumption of finite second moment of X_0^2 and the existence of m.g.f of θ_0.

Remark 36. The computations of Theorem 10 and Lemma 32 and their interpretation given in Remark 45 can easily be extended to the case when the target distribution is from a symmetric Pareto law, in particular as in $\alpha < 2$. Thus, similar conclusions may be drawn for the case when the target distribution is from a symmetric stable law.

Remark 37. From an examination of the proof of Theorem 10 it is easily seen that the finiteness of the moments of X_t results from the fact that the moments of X_0 are all finite (this is true if the starting point is non random). However we believe that if the initial data does not have a finite second ordered moment, in particular the Cauchy distribution, then the limiting distribution may be standard Cauchy.

Remark 38. For densities $\psi(\cdot)$ which has support on only one side of $\mathbb{R}$, say $\mathbb{R}^+$ one has to apply the diffusion approximation technique on $\tilde{\psi}(x) = \frac{\psi(x)+\psi(-x)}{2}$, which is the symmetric version of $\psi(\cdot)$. Then use the result here on $\tilde{\psi}$ to get samples $(X_1, X_2, \ldots, X_n)$ from $\tilde{\psi}$. Taking the absolute values $(|X_1|, |X_2|, \ldots, |X_n|)$ will give a sample from $\psi(\cdot)$.

5.3 Multi-dimensional target distribution

In this section we consider the situation when the target distribution is a multivariate distribution $\psi(\mathbf{x}),\ \mathbf{x} \in \mathbb{R}^p$. Suppose the proposal distribution is multivariate Normal $N_p(\mathbf{0}, \Sigma)$. Then the adaptive algorithm will be given as:

Algorithm 3

1. Select arbitrary $(\mathbf{X}_0, \theta_0, \xi_0) \in \mathbb{R}^p \times (0, \infty) \times \{0, 1\}$. Set $n = 1$;
2. Propose a new move, say $\mathbf{Y} \sim N_p(\mathbf{X}_{n-1}, \Sigma_{n-1})$ where $\Sigma_{n-1} = \theta_{n-1}\mathbf{I}_p$, $\mathbf{I}_p$ being the identity matrix of dimension p;
3. Accept the new point with probability $\alpha(\mathbf{X}_{n-1}, \mathbf{Y}) = \min\{1, \frac{\psi(\mathbf{Y})}{\psi(\mathbf{X}_{n-1})}\}$, $\xi_n = 1$ if the sample is accepted else $\xi_n = 0$;
4. $\theta_n = \theta_{n-1}e^{\frac{1}{\sqrt{n}}(\xi_n - q)}$, $q > 0$, $\Leftrightarrow \log(\theta_n) = \log(\theta_{n-1}) + \frac{1}{\sqrt{n}}(\xi_n - q)$;
5. $n \leftarrow n + 1$ and go to step 2.

This algorithm is equivalent to the following:

Algorithm 3′:

1. Select $\{\mathbf{X}_0, \theta_0, \xi_0\} \in \mathbb{R}^p \times (0, \infty) \times \{0, 1\}$, where $\mathbb{R}^p$ is the state space. Set $n = 1$;
2. Generate $\boldsymbol{\epsilon}_{n-1} \sim N_p(\mathbf{0}, \Sigma_{n-1})$ where $\Sigma_{n-1} = \theta_{n-1}\mathbf{I}_p$. Given $\mathbf{X}_{n-1}, \theta_{n-1}, \boldsymbol{\epsilon}_{\boldsymbol{n-1}}$ generate
$$\xi_n \sim \text{Bernoulli}\Big(\min\left\{1, \frac{\psi(\mathbf{X}_{n-1} + \theta_{n-1}\boldsymbol{\epsilon}_{\mathbf{n-1}})}{\psi(\mathbf{X}_{n-1})}\right\}\Big)$$
and then
$$\mathbf{X}_n = \mathbf{X}_{n-1} + \theta_{n-1}\xi_n\boldsymbol{\epsilon}_{n-1};$$
3. $\theta_n = \theta_{n-1}e^{\frac{1}{\sqrt{n}}(\xi_n - q)}$, $q > 0$, $\Leftrightarrow \log(\theta_n) = \log(\theta_{n-1}) + \frac{1}{\sqrt{n}}(\xi_n - q)$, $q > 0$;
4. $n \leftarrow n + 1$, and go to step 2.

REMARK 39. In Algorithm 3 all the co-ordinates $X_{in}, i = 1, \ldots, p$, for a fixed $n \geq 1$, are scaled by the same factor θ_{n-1}. This can be generalised where different co-ordinates are updated differently depending whether it is more mixing or not. We do not follow that approach here. For more information see [8].

For the multivariate AMCMC, with the multivariate Normal proposal distribution, we now state the diffusion approximation which is somewhat similar to the univariate AMCMC case as in Section 3.3 of Chapter 3. We also give the proof since it uses a slightly different method when compared to that of the univariate case and requires the spherical symmetry property of the multivariate Normal $(\mathbf{0}, \mathbf{I}_p)$ distribution.

THEOREM 11. Applying the diffusion approximation (see Section 3.3) to Algorithm 3 such that $||\nabla\psi(\mathbf{x})|| = 0$ on at most finitely many points, the diffusion corresponding to $\mathbf{Y}_t = (\mathbf{X}_t, \theta_t)$ will be the solution of the following SDE:

$$d\mathbf{Y}_t = \mathbf{b}(\mathbf{Y}_t)dt + \sigma(\mathbf{Y}_t)d\mathbf{W}_t, \tag{5.3.62}$$

where $\mathbf{b}(\mathbf{Y}_t) = \left(\frac{\theta_t^2}{2}\nabla \log\psi(\mathbf{X}_t),\ \theta_t(q - \frac{1}{\sqrt{2\pi}}\theta_t||\nabla\log\psi(\mathbf{X}_t)||)\right)^T$, and

$$\sigma(\mathbf{Y_t}) = \begin{pmatrix} \theta_t\mathbf{I}_p & \mathbf{0}_{p\times 1} \\ \mathbf{0}_{1\times p} & 0 \end{pmatrix}.$$

Here $\nabla\log\psi(\mathbf{X}_t) = \frac{1}{\psi(\mathbf{X}_t)}\left(\frac{\partial}{\partial x_{1t}}\psi(\mathbf{X}_t), \frac{\partial}{\partial x_{2t}}\psi(\mathbf{X}_t), \ldots, \frac{\partial}{\partial x_{pt}}\psi(\mathbf{X}_t)\right)^T = \frac{\nabla\psi(\mathbf{X}_t)^T}{\psi(\mathbf{X}_t)}$ is the vector of partial derivatives of $\log\psi(\mathbf{x})$, $\mathbf{X}_t = \left(X_{1t}, X_{2t}, \ldots, X_{pt}\right)^T$ is the state vector, θ_t is the tuning parameter and $\boldsymbol{W}_t = \left(W_{1t}, W_{2t}, \ldots, W_{(p+1)t}\right)^T$ is the $(p+1)$-dimensional Wiener process.

Proof: Following the arguments and notations as in Section 3.3 of Chapter 3 we have to compute the 'diffusion' and the 'drift' coefficients which in this case are defined as:

$$\begin{aligned}
\mathbf{b}_{n,1}(\mathbf{y},t) &= nE\Big(\mathbf{X}_n(\frac{i+1}{n}) - \mathbf{X}_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big), \\
\mathbf{b}_{n,2}(\mathbf{y},t) &= nE\Big(\theta_n(\frac{i+1}{n}) - \theta_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big), \\
\mathbf{A}_{n,1,1}(\mathbf{y},t) &= nE\Big((\mathbf{X}_n(\frac{i+1}{n}) - \mathbf{X}_n(\frac{i}{n}))(\mathbf{X}_n(\frac{i+1}{n}) - \mathbf{X}_n(\frac{i}{n}))^T|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big), \\
\mathbf{A}_{n,2,2}(\mathbf{y},t) &= nE\Big((\theta_n(\frac{i+1}{n}) - \theta_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big), \text{ and} \\
\mathbf{A}_{n,1,2}(\mathbf{y},t) &= nE\Big((\theta_n(\frac{i+1}{n}) - \theta_n(\frac{i}{n}))(\mathbf{X}_n(\frac{i+1}{n}) - \mathbf{X}_n(\frac{i}{n}))|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big).
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbf{b}_{n,1}(\mathbf{y},t) &= \sqrt{n}\theta\Big(E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})I_{A_n}|\ \mathbf{X}_n(\frac{i}{n}) = \mathbf{x},\ \theta_n(\frac{i}{n}) = \theta) \\
&+ E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})I_{A_n^c}|\ \mathbf{X}_n(\frac{i}{n}) = \mathbf{x},\ \theta_n(\frac{i}{n}) = \theta)\Big)
\end{aligned}$$

where $\mathcal{A}_n (= \mathcal{A}_n(\mathbf{x},\theta))$ is the set where $\xi_n(\frac{i+1}{n})$ is one with probability 1, i.e,

$$\begin{aligned}
\mathcal{A}_n &= \{\mathbf{y} : \frac{\psi(\mathbf{x} + \frac{1}{\sqrt{n}}\theta\mathbf{y})}{\psi(\mathbf{x})} \geq 1\} \\
&= \{\mathbf{y} : (\psi(\mathbf{x}) + \frac{1}{\sqrt{n}}\theta\nabla\psi(\mathbf{x})^T\mathbf{y} + O(\frac{1}{n}))/\psi(\mathbf{x}) \geq 1\} \\
&= \{\mathbf{y} : \frac{1}{\sqrt{n}}\theta\nabla\psi(\mathbf{x})^T\mathbf{y} + O(\frac{1}{n}) \geq 0\}.
\end{aligned}$$

This implies that,

$$\lim_{n\to\infty} \mathcal{A}_n = \{\mathbf{y} : \nabla\psi(\mathbf{x})^T y \geq 0\} := \mathcal{A}\ (= \mathcal{A}(\mathbf{x},\theta)).$$

Therefore,

$$\begin{aligned}
\mathbf{b}_{n,1}(\mathbf{y},t) &= \sqrt{n}\theta\Big(\int_{\mathcal{A}_n}\epsilon\phi(\epsilon)d\epsilon+\int_{\mathcal{A}_n^c}\frac{\psi(\mathbf{x}+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(\mathbf{x})}\epsilon\phi(\epsilon)d\epsilon\Big)\\
&= \sqrt{n}\theta\Big(\int_{\mathcal{A}_n}\epsilon\phi(\epsilon)d\epsilon+\int_{\mathcal{A}_n^c}(1+\frac{\theta}{\sqrt{n}\psi(\mathbf{x})}\nabla\psi(\mathbf{x})^T\epsilon)\epsilon\phi(\epsilon)+O(\frac{1}{n})\Big)\\
&= \sqrt{n}\theta\int_{\mathbb{R}^p}\epsilon\phi(\epsilon)d\epsilon+\theta^2\frac{1}{\psi(\mathbf{x})}\int_{\mathcal{A}_n^c}(\nabla\psi(\mathbf{x})^T\epsilon)\epsilon\phi(\epsilon)d\epsilon+O(\frac{1}{\sqrt{n}})\\
&= \theta^2\frac{1}{\psi(\mathbf{x})}\int_{\mathcal{A}_n^c}(\nabla\psi(\mathbf{x})^T\epsilon)\epsilon\phi(\epsilon)d\epsilon+O(\frac{1}{\sqrt{n}})\\
\Rightarrow \lim_{n\to\infty}\mathbf{b}_{n,1}(\mathbf{y},t) &= \theta^2\frac{1}{\psi(\mathbf{x})}\lim_{n\to\infty}\int_{\mathcal{A}_n^c}(\nabla\psi(\mathbf{x})^T\epsilon)\epsilon\phi(\epsilon)d\epsilon\\
&= \theta^2\frac{1}{\psi(\mathbf{x})}\int_{\mathcal{A}^c}(\nabla\psi(\mathbf{x})^T\epsilon)\ \epsilon\phi(\epsilon)d\epsilon. \qquad (5.3.63)
\end{aligned}$$

Consider the transformation

$$\epsilon :\to \mathbf{P}^T\epsilon=\mathbf{Z}=(Z_1,Z_2,\ldots,Z_p)^T, \qquad (5.3.64)$$

where $\mathbf{P}$ is an orthogonal matrix whose first column is $\frac{\nabla\psi(\mathbf{x})}{||\nabla\psi(\mathbf{x})||}:=P_1$, whenever $||\nabla\psi(\mathbf{x})||\neq 0$. Therefore $\nabla\psi(\mathbf{x})^T\epsilon=(||\nabla\psi(\mathbf{x})||P_1)^T\epsilon=||\nabla\psi(\mathbf{x})||P_1^T\epsilon=||\nabla\psi(\mathbf{x})||Z_1$. Correspondingly $\epsilon=\mathbf{PZ}$. The Jacobian of the transformation (5.3.64) is 1 and $Z_i, i=1,2,\ldots,p$ are i.i.d $N(0,1)$, since $\epsilon_i,\ i=1,2,\ldots,p$ are also i.i.d standard Normal. The integral in the RHS of Equation (5.3.63) is therefore

$$\begin{aligned}
\int_{\mathcal{A}^c}(\nabla\psi(\mathbf{x})^T\epsilon)\epsilon\phi(\epsilon)d\epsilon &= ||\nabla\psi(\mathbf{x})||\int_{\{Z_1<0\}}Z_1\mathbf{PZ}\phi(\mathbf{z})d\mathbf{z}\\
&= ||\nabla\psi(\mathbf{x})||\int_{\{Z_1<0\}}Z_1\sum_{i=1}^{p}P_iZ_i\phi(\mathbf{z})d\mathbf{z}\\
&= ||\nabla\psi(\mathbf{x})||\Big(P_1\int_{\{Z_1<0\}}Z_1^2\phi(\mathbf{z})d\mathbf{z}+\sum_{i=2}^{p}P_i\int_{\{Z_1<0\}}Z_1Z_i\phi(\mathbf{z})d\mathbf{z}\Big)\\
&= ||\nabla\psi(\mathbf{x})||P_1\int_{\{Z_1<0\}}Z_1^2\phi(\mathbf{z})d\mathbf{z}\ \text{(since } Z_i\text{'s are independent and } E(Z_i)=0)\\
&= ||\nabla\psi(\mathbf{x})||\frac{P_1}{2}\\
&= \frac{1}{2}\nabla\psi(\mathbf{x}),\ \text{since } P_1=\tfrac{\nabla\psi(\mathbf{x})}{||\nabla\psi(\mathbf{x})||}.
\end{aligned}$$

Therefore

$$\mathbf{b}_1(\mathbf{y},t) \;=\; \frac{\theta^2}{2}\frac{\nabla\psi(\mathbf{x})}{\psi(\mathbf{x})} = \frac{\theta^2}{2}\nabla\log\psi(\mathbf{x}). \tag{5.3.65}$$

For $b_{n,2}(\mathbf{y},t)$ we have

$$\begin{aligned}
b_{n,2}(\mathbf{y},t) &= nE(\theta_n(\frac{i+1}{n}) - \theta_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}), \quad \forall i = 0,1,\ldots \\
&= nE\Big(\theta_n(\frac{i}{n})\{e^{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))} - 1\}|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
&= n\theta\Big(\frac{1}{\sqrt{n}}E(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) \\
&+ E(\frac{1}{2n}(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) + O(\frac{1}{n^{3/2}})\Big) \\
&= \theta\sqrt{n}E(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) \\
&+ \frac{\theta}{2}E((\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) + O(\frac{1}{\sqrt{n}}).
\end{aligned}$$

Now for the first term ,

$$\begin{aligned}
&\theta\sqrt{n}E(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) \\
= \;&\theta\sqrt{n}\Big(E(\xi_n(\frac{i+1}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) - q_n(\frac{i}{n})\Big) \\
= \;&\theta\sqrt{n}\Big(\int_{A_n}\phi(\epsilon)d\epsilon + \int_{A_n^c}\frac{\psi(\mathbf{x}+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(\mathbf{x})}\phi(\epsilon)d\epsilon - q_n(\frac{i}{n})\Big) \\
= \;&\theta\sqrt{n}\Big(\int_{\mathcal{A}_n}\phi(\epsilon)d\epsilon + \int_{\mathcal{A}_n^c}\{1 + \frac{\theta}{\sqrt{n}}\frac{\nabla\psi(\mathbf{x})^T}{\psi(\mathbf{x})}\epsilon + O(\frac{1}{n})\}\phi(\epsilon)d\epsilon - q_n(\frac{i}{n})\Big) \\
= \;&\theta\sqrt{n}(1 - q_n(\frac{i}{n})) + \frac{\theta^2}{\psi(\mathbf{x})}\int_{\mathcal{A}_n^c}\nabla\psi(\mathbf{x})^T\epsilon\phi(\epsilon)d\epsilon + O(\frac{1}{\sqrt{n}}).
\end{aligned}$$

And for the second term ,

$$\begin{aligned}
& E\Big((\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
= \;& E\Big(\xi_n(\frac{i+1}{n})^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
- \;& 2q_n(\frac{i}{n})E\Big(\xi_n(\frac{i+1}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) + q_n(\frac{i}{n})^2 \\
= \;& \int_{\mathcal{A}_n} \phi(\epsilon)d\epsilon + \int_{\mathcal{A}_n^c} \frac{\psi(\mathbf{x} + \frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(\mathbf{x})}\phi(\epsilon)d\epsilon \\
- \;& 2q_n(\frac{i}{n})\Big(\int_{\mathcal{A}_n} \phi(\epsilon)d\epsilon + \int_{\mathcal{A}_n^c} \frac{\psi(\mathbf{x} + \frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(\mathbf{x})}\phi(\epsilon)d\epsilon\Big) + q_n(\frac{i}{n})^2 \\
= \;& (1 - q_n(\frac{i}{n}))^2 + \frac{1}{\sqrt{n}}(1 - 2q_n(\frac{i}{n}))\theta\frac{1}{\psi(\mathbf{x})}\int_{\mathcal{A}_n^c} \nabla\psi(\mathbf{x})^T\epsilon\phi(\epsilon)d\epsilon + O(\frac{1}{n}) \\
\to \;& 0,
\end{aligned}$$

as $n \to \infty$, since as before we assume that $1 - q_n(\frac{i}{n}) \approx \frac{q}{\sqrt{n}}$. Therefore

$$\frac{1}{\sqrt{n}}(1 - 2q_n(\frac{i}{n})) \;\approx\; \frac{1}{\sqrt{n}}(\frac{2q}{\sqrt{n}} - 1).$$

Thus, only the first term contributes and we have

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{b}_{n,2}(\mathbf{y}, t) &= \theta q + \frac{\theta^2}{\psi(\mathbf{x})}\lim_{n\to\infty}\int_{\mathcal{A}_n^c} (\nabla\psi(\mathbf{x})^T\epsilon)\phi(\epsilon)d\epsilon \\
&= \theta q + \frac{\theta^2}{\psi(\mathbf{x})}\int_{\mathcal{A}^c} (\nabla\psi(\mathbf{x})^T\epsilon)\phi(\epsilon)d\epsilon. \qquad (5.3.66)
\end{aligned}$$

Using the transformation used in Equation (5.3.64) above we have

$$\begin{aligned}
\int_{\mathcal{A}^c} (\nabla\psi(\mathbf{x})^T)\epsilon\phi(\epsilon)d\epsilon &= ||\nabla\psi(\mathbf{x})||\int_{\{Z_1<0\}} Z_1\phi(\mathbf{z})d\mathbf{z} \\
&= ||\nabla\psi(\mathbf{x})||E(Z_1 I(Z_1 < 0)) \\
&= -\frac{1}{\sqrt{2\pi}}||\nabla\psi(\mathbf{x})||.
\end{aligned}$$

Therefore from (5.3.66) we have

$$b_2(\mathbf{y}, t) = \theta\Big(q - \frac{1}{\sqrt{2\pi}}\frac{||\nabla\psi(\mathbf{x})||}{\psi(\mathbf{x})}\theta\Big) = \theta\Big(q - \frac{1}{\sqrt{2\pi}}\theta||\nabla\log\psi(\mathbf{x})||\Big)$$

$$\begin{aligned}
\mathbf{A}_{n,1,1}(\mathbf{y},t) &= nE\Big((\mathbf{X}_n(\frac{i+1}{n})-\mathbf{X}_n(\frac{i}{n}))(\mathbf{X}_n(\frac{i+1}{n})-\mathbf{X}_n(\frac{i}{n}))^T|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\ \forall i=0,1,\ldots\\
&= \theta^2E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})^T|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y})\\
&= \theta^2\Big(E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})^TI_{A_n}|\ \mathbf{Y}_n(\frac{i}{n})=\mathbf{y})\\
&+ E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})^TI_{A_n^c}|\ \mathbf{Y}_n(\frac{i}{n})=\mathbf{y})\Big)\\
&= \theta^2\Big(\int_{A_n}\epsilon\epsilon^T\phi(\epsilon)d\epsilon+\int_{A_n^c}\epsilon\epsilon^T\frac{\psi(\mathbf{x}+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(\mathbf{x})}\phi(\epsilon)d\epsilon\Big)\\
&= \theta^2\Big(\int_{A_n}\epsilon\epsilon^T\phi(\epsilon)d\epsilon+\int_{A_n^c}\epsilon\epsilon^T\phi(\epsilon)d\epsilon+O(\frac{1}{\sqrt{n}})\Big)\\
&= \theta^2\int_{\mathbb{R}^p}\epsilon\epsilon^T\phi(\epsilon)d\epsilon+O(\frac{1}{\sqrt{n}})=\theta^2\mathbf{I}_p+O(\frac{1}{\sqrt{n}}).\\
\Rightarrow \lim_{n\to\infty}\mathbf{A}_{n,1,1}(\mathbf{y},t) &= \theta^2\mathbf{I}_p.
\end{aligned}$$

The computations for $\mathbf{A}_{2,2}(\mathbf{y},t)$ is same as that of the univariate case and is not repeated here.

$$\begin{aligned}
\mathbf{A}_{n,1,2}(\mathbf{y},t) &= nE\Big(\{\mathbf{X}_n(\frac{i+1}{n})-\mathbf{X}_n(\frac{i}{n})\}\{\theta_n(\frac{i+1}{n})-\theta_n(\frac{i}{n})\}|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&= nE\Big(\{\frac{1}{\sqrt{n}}\theta_n(\frac{i}{n})\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})\}\{\theta_n(\frac{i}{n})(e^{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))}-1)\}|\mathbf{Y}_n(\frac{1}{n})=\mathbf{y}\Big)\\
&= \sqrt{n}\theta^2E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})\Big\{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))+O(\frac{1}{n})\Big\}|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&= \theta^2E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&+ O(\frac{1}{\sqrt{n}}).
\end{aligned}$$

Since $\xi_n=0$, or 1, $\xi_n^2=\xi_n$. Hence $\xi_n\epsilon_{\mathbf{n}}(\xi_n-q_n)=\xi_n^2\epsilon_n-\xi_n\epsilon_nq_n=\xi_n\epsilon_n(1-q_n)$. Therefore,

$$\begin{aligned}
&E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&= (1-q_n(\frac{i}{n}))E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big)\\
&= (1-q_n(\frac{i}{n}))O(1)\ \longrightarrow 0,\quad \text{as}\quad n\to\infty.
\end{aligned}$$

Similar computation for $\mathbf{A}_{n,2,1}$ yields that $\lim_{n\to\infty} \mathbf{A}_{n,2,1} = \lim_{n\to\infty} \mathbf{A}_{n,1,2} = 0$. The drift and the diffusion coefficient gives the limiting diffusion of the multivariate AMCMC and this proves the theorem. ■

REMARK 40. For the uniqueness and the non-explosion of the solutions of the above SDE (5.3.62) we need the local Lipschitz and the linear growth conditions as in Remark 12 and 13 of Chapter 3. This in particular would mean that $\nabla \log \psi(\cdot)$ satisfies the linear growth condition

$$||\nabla \log \psi(\mathbf{x})|| \quad \leq \quad a||\mathbf{x}|| + b, \; \forall \mathbf{x} \in \mathbb{R}^p, \tag{5.3.67}$$

for some $a > 0$ and $b \geq 0$.

REMARK 41. Following the arguments similar to the proof of Theorem 7 and Lemma 25 for proving the boundedness of the moments of X_t and η_t respectively, one can give similar bounds for $\sup\limits_{t>0} E(||X_t||^{2k})$ and $\sup\limits_{t>0} E(\eta_t^{2k})$ under the assumption

$$x^T \frac{\nabla \psi(x)}{\psi(x)} \leq -c||x||^2 + d,$$

for some $c > 0$ and $d \geq 0$ and the linear growth condition (5.3.67). Thus we get the tightness of the $p+1$ dimensional process $\{\mathbf{X}_t, \eta_t\}$. Arguing similarly as in Section 4.2.2, one can show that the hypoelliptic conditions are satisfied. Hence, the limiting density of $\{\mathbf{X}_t, \eta_t\}$ exists. Using the argument as in Section 4.2.3 of Chapter 4 one can identify the X-marginal of the process when the target distribution is Multivariate Normal.

5.4 Stein's lemma for Standard MCMC

Consider the Stein's identity for a density $\psi(\cdot)$ (not necessarily standard Normal.) Let X be a random variable having density π and $\psi(\cdot)$ be any other density such that

$$E_\pi\Big(f'(X)\frac{\psi'(X)}{\psi(X)}\Big) + E_\pi\Big(f''(X)\Big) = 0, \tag{5.4.68}$$

where f belongs to the class of functions for which the second derivatives exist and the expectations $E(f'(X)\frac{\psi'(X)}{\psi(X)})$ and $E(f''(X))$ is finite, see, for example, [3, 16], then $\pi(\cdot)$ is the same as $\psi(\cdot)$.

Recall that the SDE corresponding to the standard MCMC (where the tuning parameter θ_t is a constant θ_0) with the target density $\psi(\cdot)$ is given by

$$dX_t = \frac{\theta_0^2}{2}\frac{\psi'(X_t)}{\psi(X_t)}dt + \theta_0 dW_t,$$

where $\theta_0 > 0$. Let $\pi(\cdot)$ be the stationary density of $\{X_t\}$. Now applying the Itô's lemma to a bounded and twice continuously differentiable function $f(\cdot)$ we get

$$\begin{aligned}
f(X_t) - f(X_0) &= \frac{\theta_0^2}{2}\int_0^t \Big(f'(X_s)\frac{\psi'(X_s)}{\psi(X_s)} + f''(X_s)\Big)ds + \int_0^t \theta_0 f'(X_s)dW_s \\
\Rightarrow \frac{E(f(X_t)) - E(f(X_0))}{t} &= \frac{\theta_0^2}{2}\frac{1}{t}\Big(E\int_0^t \Big(f'(X_s)\frac{\psi'(X_s)}{\psi(X_s)} + f''(X_s)\Big)ds + E\int_0^t \theta_0 f'(X_s)dW_s\Big) \\
&= \frac{\theta_0^2}{2}\frac{1}{t}\int_0^t E\Big(f'(X_s)\frac{\psi'(X_s)}{\psi(X_s)} + f''(X_s)\Big)ds \\
\Rightarrow \lim_{t\uparrow\infty}\frac{E(f(X_t)) - E(f(X_0))}{t} &= \frac{\theta_0^2}{2}\lim_{t\uparrow\infty}\frac{1}{t}\int_0^t E\Big(f'(X_s)\frac{\psi'(X_s)}{\psi(X_s)} + f''(X_s)\Big)ds. \qquad (5.4.69)
\end{aligned}$$

If the SMCMC is ergodic, see definition in Chapter 4.1, then the time average converges to state average almost surely. This means that for a function $f(\cdot)$ belonging to a suitable class

$$\lim_{t\uparrow\infty}\frac{1}{t}\int_0^t \Big(f'(X_s)\frac{\psi'(X_s)}{\psi(X_s)}\Big) = E_\pi(f'(X)\frac{\psi'(X)}{\psi(X)}). \qquad (5.4.70)$$

Now if moments of $f(X_t)$ are uniformly bounded in $t > 0$ then using the fact that X_s admits a limiting distribution $\pi(\cdot)$ and by applying Equation (5.4.70) and L_1 convergence for a suitable class of functions f to the RHS of the last equality we get,

$$0 = E_\pi\Big(f'(X)\frac{\psi'(X)}{\psi(X)}\Big) + E_\pi\Big(f''(X)\Big)$$

Therefore it satisfies the Stein's identity in Equation (5.4.68) and therefore $\pi = \psi$. In particular, when $\psi(x)$ is the standard Normal density then $\frac{\psi'(x)}{\psi(x)} = -x$ and therefore the

Stein's identity is

$$E(f''(X)) - E(Xf'(X)) \;\;=\;\; 0.$$

Writing $g = f'$ the above equation takes the more familiar form

$$E(g'(X)) - E(Xg(X)) \;\;=\;\; 0.$$

REMARK 42. For the AMCMC Equation (5.4.69) takes the form

$$0 \;\;=\;\; \lim_{t\uparrow\infty} \frac{1}{t} \int_0^t E\Big(f'(X_s)\theta_s^2 \frac{\psi'(X_s)}{\psi(X_s)} + f''(X_s)\theta_s^2\Big) ds.$$

If the moments of X_t are uniformly bounded in $t > 0$ then applying the ergodic theorem and the L_1 convergence for suitably chosen f we have

$$0 \;\;=\;\; E(f'(X)\frac{\psi'(X)}{\psi(X)}\theta^2) + E(f''(X)\theta^2),$$

where the expectation is taken with respect to the joint limiting distribution of X and θ. We believe that it can be shown that only the marginal of X will also satisfy the Stein's identity. We plan to take this up as a future research. Also see the Remark at the end of Section 5.5.

5.5 Further heavy tailed target densities

In this section we consider an extension to Section 5.2.4 where the target density is of the form

$$\psi(x) \propto \frac{1}{1 + x^{2k}}, \;\; x \in \mathbb{R}, \;\; k \in \mathbb{N}. \tag{5.5.71}$$

This corresponds to a heavy tailed target distribution whose all moments smaller than $2k - 1$ exists.

THEOREM 12. If $\psi(\cdot)$ is the density given by Equation (5.5.71) and $E(X_0^{2k}) < \infty$ then $\limsup_{t\to\infty} E(X_t^{2k}) < \infty$ where k is as in Equation (5.5.71).

Proof: The proof will emulate the proof of Theorem 10. First note that the SDE corresponding to X_t is:

$$dX_t = -\frac{kX_t^{2k-1}\theta_t^2}{1+X_t^{2k}}dt + \theta_t dW_t.$$

Let $V_t = 1 + X_t^{2k}$. Then the SDE corresponding to V_t will

$$\begin{aligned}
dV_t &= 2kX_t^{2k-1}dX_t + k(2k-1)X_t^{2k-2}\theta_t^2 dt \\
&= 2kX_t^{2k-1}\Big(-\frac{kX_t^{2k-1}\theta_t^2}{1+X_t^{2k}}dt + \theta_t dW_t\Big) + k(2k-1)X_t^{2k-2}\theta_t^2 dt \\
&= \frac{k\theta_t^2\Big(-2kX_t^{4k-2} + (2k-1)X_t^{2k-2} + (2k-1)X_t^{4k-2}\Big)}{1+X_t^{2k}}dt + 2kX_t^{2k-1}\theta_t dW_t \\
&= \frac{k\theta_t^2(-X_t^{4k-2} + (2k-1)X_t^{2k-2})}{1+X_t^{2k}}dt + 2kX_t^{2k-1}\theta_t dW_t \\
&= \frac{k\theta_t^2\Big(-X_t^{2k-2}(1+X_t^{2k}) + 2kX_t^{2k-2}\Big)}{1+X_t^{2k}}dt + 2kX_t^{2k-1}\theta_t dW_t \\
&= -k\theta_t^2X_t^{2k-2}dt + 2k^2\frac{\theta_t^2X_t^{2k-2}}{1+X_t^{2k}}dt + 2kX_t^{2k-1}\theta_t dW_t \\
\Rightarrow dV_t + k\theta_t^2X_t^{2k-2}dt &= 2k^2\theta_t^2X_t^{2k-2}V_t dt + 2kX_t^{2k-1}\theta_t dW_t.
\end{aligned}$$

Multiplying by the integrating factor $e^{k\int_0^t \theta_s^2 X_s^{2k-2}V_s^{-1}ds} := e^{C_t}$ on both sides we have

$$d\Big(V_te_t^C\Big) = 2ke^{C_t}k\theta_t^2X_t^{2k-2}V_t dt + 2ke^{C_t}X_t^{2k-1}\theta_t dW_t.$$

Integrating from 0 to t gives

$$\begin{aligned}
V_te^{C_t} - V_0 &= 2k\int_0^t e^{C_s}k\theta_s^2X_s^{2k-2}V_s ds + 2k\int_0^t e^{C_s}X_s^{2k-1}\theta_s dW_s \\
&= 2k(e^{C_t}-1) + 2k\int_0^t e^{C_s}X_s^{2k-1}\theta_s dW_s.
\end{aligned}$$

By arguing similarly as in Lemma 31 we have that

$$E(e^{-C_t}\int_0^t e^{C_s}X_s^{2k-1}\theta_s ds) = 0.$$

Hence taking expectations on both sides we have

$$\begin{aligned} E(V_t) &= E(V_0 e^{-C_t}) + 2k(1 - E(e^{-C_t})) \\ \Rightarrow \limsup_{t\to\infty} E(X_t^{2k}) &< \infty, \end{aligned}$$

since $e^{-C_t} \leq 1$ almost surely and $E(V_0) < \infty$ by the hypothesis of the theorem. ■

REMARK 43. Using the Echeverria's theorem (see Theorem 9.17, Chapter 4 of [22]) it might be possible to identify the invariant distribution for the diffusion corresponding to SMCMC or AMCMC. However the problem lies not in identifying the invariant distribution but the limiting distribution, since using SMCMC or AMCMC it is the limiting distribution from which samples are generated. Theorem 10 and 12 proves that the limiting distribution does not match with the target $\psi(\cdot)$. Hence AMCMC or SMCMC should not be used in simulating from such Pareto type densities which admits only a finitely many moments. We feel that only in those distributions where all the moments exist (or in particular, the m.g.f exists in a neighborhood of zero) the limiting distribution of the diffusion may coincide with the invariant distribution which is the target distribution. Hence only in those cases should the SMCMC or AMCMC be applied for simulating the target distribution.

5.6 Choice of the proposal distribution

There is nothing special about the Normal distribution as the choice of the proposal distribution in the univariate case. In fact, any distribution whose support is $\mathbb{R}$, symmetric about its mean and has finite variance will also yield similar results.

THEOREM 13. Consider a RW MH algorithm where the proposal density is given by $q(x, y) = f(y - x)$ where $f(\cdot)$ is a density with finite variance σ^2. Then the SDE corresponding to $\{X_t\}$ is :

$$\begin{aligned} dX_t &= \frac{\sigma^2\theta_t^2}{2}\frac{\psi'(X_t)}{\psi(X_t)}dt + \sigma\theta_t dW_t; \\ d\theta_t &= \theta_t\Big(q - c\theta_t\frac{|\psi'(X_t)|}{\psi(X_t)}\Big)dt, \end{aligned}$$

where $c = \int_0^\infty \epsilon f(\epsilon)d\epsilon \in (0,\infty)$ is the one sided mean of the density $f(\cdot)$.

Proof: The diffusion approximation in this case is similar to the diffusion approximation as in Section 3.3.1 except that now the proposal will be a general distribution $f(\cdot)$:

$$\begin{aligned} X_n(0) &= x_0 \in \mathbb{R}; \\ X_n(\frac{i+1}{n}) &= X_n(\frac{i}{n}) + \frac{1}{\sqrt{n}}\theta_n(\frac{i}{n})\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n}),\ i = 0,1,\dots \\ X_n(t) &= X_n(\frac{i}{n}),\ \text{if } \frac{i}{n} \le t < \frac{i+1}{n} \text{ for some integer } i. \end{aligned} \tag{5.6.72}$$

Here, $\xi_n(\frac{i}{n})$ conditionally follows the Bernoulli distribution with parameter $p \in (0,1)$ given by

$$\min\{\frac{\psi(X_n(\frac{i}{n}) + \frac{1}{\sqrt{n}}\theta_n(\frac{i}{n})\epsilon_n(\frac{i}{n}))}{\psi(X_n(\frac{i}{n}))}, 1\},$$

where $\epsilon_n(\frac{i}{n})$ is a random variable having density $f(\cdot)$. Hence, the The proof will emulate the proof when the proposal is standard Normal. Then as in Section 3.5 we compute the constants $a's$ and $b's$. Writing $\mathbf{y} = (x,\theta)$ we have for every fixed $n \ge 1$ we define

$$\begin{aligned} & b_{n,1}(\mathbf{y},t) \\ = \ & nE(X_n(\frac{i+1}{n}) - X_n(\frac{i}{n})|\ \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}), \quad \forall i = 0,1,\dots, \\ = \ & E(\sqrt{n}\theta_n(\frac{i}{n})\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})|\ \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) \\ = \ & \sqrt{n}\theta\Big(E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})I_{A_n}|\ X_n(\frac{i}{n}) = x,\ \theta_n(\frac{i}{n}) = \theta) \\ + \ & E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})I_{A_n^c}|\ X_n(\frac{i}{n}) = x,\ \theta_n(\frac{i}{n}) = \theta)\Big). \end{aligned}$$

where $A_n (= A_n(x,\theta))$ is the set where $\xi_n(\frac{i+1}{n})$ is one with probability 1, i.e,

$$A_n(x,\theta) \;=\; \{y : \frac{\psi(x+\frac{1}{\sqrt{n}}\theta y)}{\psi(x)} \geq 1\}.$$

$$\text{Thus,} \quad \lim_{n\to\infty} A_n^c(x,\theta) \;=\; \begin{cases} (-\infty,0) & \text{if } \psi'(x) > 0 \\ (0,\infty) & \text{if } \psi'(x) < 0. \end{cases}$$

Therefore,

$$\begin{aligned}
b_{n,1}(\mathbf{y},t) &= \sqrt{n}\theta\Big(\int_{A_n} \epsilon f(\epsilon)d\epsilon + \int_{A_n^c} \frac{\psi(x+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)}\epsilon f(\epsilon)d\epsilon\Big) \\
&= \sqrt{n}\theta\Big(\int_{A_n} \epsilon f(\epsilon)d\epsilon + \int_{A_n^c} \epsilon f(\epsilon)d\epsilon \\
&+ \frac{\theta}{\sqrt{n}}\frac{\psi'(x)}{\psi(x)}\int_{A_n^c} \epsilon^2 f(\epsilon)d\epsilon + O(\frac{1}{n})\Big), \quad \text{by Taylor's expansion,} \\
&= \sqrt{n}\theta\Big(\int_{\mathbb{R}} \epsilon f(\epsilon)d\epsilon + \frac{\theta}{\sqrt{n}}\frac{\psi'(x)}{\psi(x)}\int_{A_n^c} \epsilon^2 f(\epsilon)d\epsilon + O(\frac{1}{n})\Big) \\
&= \theta^2\frac{\psi'(x)}{\psi(x)}\int_{A_n^c} \epsilon^2 f(\epsilon)d\epsilon + O(\frac{1}{\sqrt{n}}) \\
\Rightarrow \lim_{n\to\infty} b_{n,1}(\mathbf{y},t) &= \theta^2\frac{\psi'(x)}{\psi(x)}\lim_{n\to\infty}\int_{A_n^c} \epsilon^2 f(\epsilon)d\epsilon \\
&= \begin{cases} \theta^2\frac{\psi'(x)}{\psi(x)}\int_{-\infty}^0 \epsilon^2 f(\epsilon)d\epsilon & \text{if } \psi'(x) > 0 \\ \theta^2\frac{\psi'(x)}{\psi(x)}\int_0^\infty \epsilon^2 f(\epsilon)d\epsilon & \text{if } \psi'(x) < 0 \end{cases} \\
&= \frac{\sigma^2\theta^2}{2}\frac{\psi'(x)}{\psi(x)},
\end{aligned}$$

since the distribution is symmetric about 0 and the variance is $\sigma^2 = 2\int_0^\infty \epsilon^2 f(\epsilon)d\epsilon$.

$$
\begin{aligned}
b_{n,2}(\mathbf{y},t) &= nE(\theta_n(\frac{i+1}{n})-\theta_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}), \quad \forall i=0,1,\ldots \\
&= nE\Big(\theta_n(\frac{i}{n})\{e^{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))}-1\}|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}\Big) \\
&= n\theta\Big(\frac{1}{\sqrt{n}}E(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}) \\
&+ E(\frac{1}{2n}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y})+O(\frac{1}{n^{3/2}})\Big) \\
&= \theta\sqrt{n}E(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}) \\
&+ \frac{\theta}{2}E((\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))^2)|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y})+O(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\theta\sqrt{n}E(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n})|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y}) \\
=\ & \theta\sqrt{n}\Big(E(\xi_n(\frac{i+1}{n})|\mathbf{Y}_n(\frac{i}{n})=\mathbf{y})-q_n(\frac{i}{n})\Big) \\
=\ & \theta\sqrt{n}\Big(\int_{A_n}f(\epsilon)d\epsilon+\int_{A_n^c}\frac{\psi(x+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)}f(\epsilon)d\epsilon-q_n(\frac{i}{n})\Big) \\
=\ & \theta\sqrt{n}\Big(\int_{A_n}f(\epsilon)d\epsilon \\
+\ & \int_{A_n^c}\{1+\frac{\theta}{\sqrt{n}}\frac{\psi'(x)}{\psi(x)}\epsilon+O(\frac{1}{n})\}f(\epsilon)d\epsilon-q_n(\frac{i}{n})\Big) \\
=\ & \theta\sqrt{n}(1-q_n(\frac{i}{n})) \\
+\ & \theta^2\frac{\psi'(x)}{\psi(x)}\int_{A_n^c}\epsilon f(\epsilon)d\epsilon+O(\frac{1}{\sqrt{n}}).
\end{aligned} \tag{5.6.73}
$$

And,

$$\begin{aligned}
& E\Big((\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))^2 | \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
= \;& E\Big(\xi_n(\frac{i+1}{n})^2 | \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
- \;& 2q_n(\frac{i}{n}) E\Big(\xi_n(\frac{i+1}{n}) | \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) + q_n(\frac{i}{n})^2 \\
= \;& \int_{A_n} f(\epsilon)d\epsilon + \int_{A_n^c} \frac{\psi(x + \frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)} f(\epsilon)d\epsilon \\
- \;& 2q_n(\frac{i}{n}) \Big(\int_{A_n} f(\epsilon)d\epsilon + \int_{A_n^c} \frac{\psi(x + \frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)} f(\epsilon)d\epsilon\Big) + q_n(\frac{i}{n})^2 \\
= \;& (1 - q_n(\frac{i}{n}))^2 + \frac{1}{\sqrt{n}}(1 - 2q_n(\frac{i}{n}))\theta \frac{\psi'(x)}{\psi(x)} \int_{A_n^c} \epsilon f(\epsilon)d\epsilon + O(\frac{1}{n}) \\
\longrightarrow \;& 0, \quad \text{as } \; n \to \infty, \quad \text{since } 1 - q_n(\frac{i}{n}) \approx \frac{q}{\sqrt{n}}, \quad \text{and hence,}
\end{aligned}$$

$$\frac{1}{\sqrt{n}}(1 - 2q_n(\frac{i}{n})) \approx \frac{1}{\sqrt{n}}(\frac{2q}{\sqrt{n}} - 1). \tag{5.6.74}$$

Thus, from (5.6.73) and (5.6.74) we have,

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{b}_{n,2}(\mathbf{y}, t) &= \theta q + \theta^2 \frac{\psi'(x)}{\psi(x)} \lim_{n\to\infty} \int_{A_n^c} \epsilon f(\epsilon)d\epsilon \\
&= \begin{cases} \theta\Big(q + c\theta\frac{\psi'(x)}{\psi(x)}\Big) & \text{if } \psi'(x) < 0 \\ \theta\Big(q - c\theta\frac{\psi'(x)}{\psi(x)}\Big) & \text{if } \psi'(x) > 0 \end{cases} \\
&= \theta\Big(q - c\theta\frac{|\psi'(x)|}{\psi(x)}\Big),
\end{aligned}$$

where c is the one sided mean of $f(\cdot)$ given by $c = \int_0^\infty \epsilon f(\epsilon)d\epsilon = -\int_{-\infty}^0 \epsilon f(\epsilon)d\epsilon < \infty$.

$$
\begin{aligned}
a_{n,1,1}(\mathbf{y},t) &= nE\Big((X_n(\frac{i+1}{n}) - X_n(\frac{i}{n})^2)|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \ \ \forall i = 0, 1, \ldots \\
&= \theta^2 E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) \\
&= \theta^2\Big(E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})^2 I_{A_n}|\ \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}) \\
&+ E(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})^2 I_{A_n^c}|\ \mathbf{Y}_n(\frac{i}{n}) = \mathbf{y})\Big) \\
&= \theta^2\Big(\int_{A_n} \epsilon^2 f(\epsilon)d\epsilon + \int_{A_n^c} \frac{\psi(x+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)}\epsilon^2 f(\epsilon)d\epsilon\Big) \\
&= \theta^2\Big(\int_{A_n} \epsilon^2 f(\epsilon)d\epsilon + \int_{A_n^c} \epsilon^2 f(\epsilon)d\epsilon + O(\frac{1}{\sqrt{n}})\Big) \\
&= \sigma^2\theta^2 + O(\frac{1}{\sqrt{n}}). \\
\Rightarrow \lim_{n\to\infty} a_{n,1,1}(t) &= \sigma^2\theta^2.
\end{aligned}
$$

$$
\begin{aligned}
a_{n,2,2}(\mathbf{y},t) &= nE\Big((\theta_n(\frac{i+1}{n}) - \theta_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
&= nE\Big(\theta_n(\frac{i}{n})^2(e^{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))} - 1)^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
&= n\theta^2 E\Big(\Big\{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n})) + \frac{1}{2n}(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))^2 \\
&+ O(\frac{1}{n^{3/2}})\Big\}^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
&= \theta^2 E\Big((\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) + O(\frac{1}{\sqrt{n}}) \\
\Rightarrow \lim_{n\to\infty} a_{n,2,2}(\mathbf{y},t) &= \theta^2 \lim_{n\to\infty} E\Big(\Big(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n})\Big)^2|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) = 0, \quad \text{from (5.6.74)}.
\end{aligned}
$$

$$\begin{aligned}
a_{n,1,2}(\mathbf{y},t) &= nE\Big(\{X_n(\frac{i+1}{n}) - X_n(\frac{i}{n})\}\{\theta_n(\frac{i+1}{n}) - \theta_n(\frac{i}{n})\}|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
&= nE\Big(\{\frac{1}{\sqrt{n}}\theta_n(\frac{i}{n})\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})\}\{\theta_n(\frac{i}{n})(e^{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n})-q_n(\frac{i}{n}))})\}\Big) \\
&= \sqrt{n}\theta^2 E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})\Big\{\frac{1}{\sqrt{n}}(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n})) + O(\frac{1}{n})\Big\}|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
&= \theta^2 E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) + O(\frac{1}{\sqrt{n}}).
\end{aligned}$$

Since $\xi_n = 0$, or 1, $\xi_n^2 = \xi_n$. Hence $\xi_n\epsilon_n(\xi_n - q_n) = \xi_n^2\epsilon_n - \xi_n\epsilon_n q_n = \xi_n\epsilon_n(1 - q_n)$. Therefore,

$$\begin{aligned}
&E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})(\xi_n(\frac{i+1}{n}) - q_n(\frac{i}{n}))|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
&= (1 - q_n(\frac{i}{n}))E\Big(\xi_n(\frac{i+1}{n})\epsilon_n(\frac{i+1}{n})|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big) \\
&= (1 - q_n(\frac{i}{n}))O(1) \longrightarrow 0, \quad \text{as} \quad n \to \infty.
\end{aligned}$$

Thus, $\lim_{n\to\infty} a_{n,1,2} = \lim_{n\to\infty} a_{n,2,1} = 0$.
This proves the theorem. ■

Example 3: The t distribution with ν degrees of freedom is symmetric about 0 and has finite mean and variance equal to $\frac{\nu}{\nu-2}$ for $\nu > 2$. So if we choose the t distribution with 3 degrees of freedom as the proposal distribution then :

$$\begin{aligned}
c &= \int_0^\infty xf(x)dx \\
&= \frac{\Gamma(2)}{\sqrt{3\pi}\Gamma(3/2)}\int_0^\infty x\Big(1 + \frac{x^2}{3}\Big)^{-2}dx = \frac{2}{\sqrt{3}\pi}\int_0^\infty x\Big(1 + \frac{x^2}{3}\Big)^{-2}dx \\
&= \frac{\sqrt{3}}{\pi}\int_1^\infty z^{-2}dz = \frac{\sqrt{3}}{\pi}.
\end{aligned}$$

For the t distribution with $\nu = 3, \sigma^2 = 3$. Therefore the diffusion would satisfy the SDE:

$$\begin{aligned} dX_t &= \frac{3\theta_t^2}{2}\frac{\psi'(X_t)}{\psi(X_t)}dt + \sqrt{3}\theta_t dW_t, \\ d\theta_t &= \theta_t\Big(q - \frac{\sqrt{3}}{\pi}\theta_t\frac{|\psi'(X_t)|}{\psi(X_t)}\Big)dt. \end{aligned}$$

REMARK 44. In fact it is not even necessary that the proposal distribution be symmetric about the mean. Suppose that the mean is zero and $c = \int_0^\infty \epsilon f(\epsilon)d\epsilon$. Since the mean is zero, we have that, $-c = \int_{-\infty}^0 \epsilon f(\epsilon)d\epsilon$. Let the variance of the proposal distribution be σ^2. Let $a = \int_0^\infty \epsilon^2 f(\epsilon)d\epsilon = \sigma^2 - b = \sigma^2 - \int_{-\infty}^0 \epsilon^2 f(\epsilon)d\epsilon > 0$. Then following the proof of the above theorem one finds that the limiting diffusion will be the solution of the SDE:

$$\begin{aligned} dX_t &= \frac{1}{\sigma^2}\Big(aI(\psi'(X_t) < 0) + bI(\psi'(X_t) \geq 0)\Big)\ \theta_t^2\frac{\psi'(X_t)}{\psi(X_t)}dt + \sigma\theta_t dW_t, \\ d\theta_t &= \theta_t\Big(q - c\theta_t\frac{|\psi'(X_t)|}{\psi(X_t)}\Big)dt. \end{aligned}$$

5.6.1 Standard Cauchy as proposal

We show that if the Cauchy distribution is chosen as the proposal distribution then the infinitesimal restricted drift (see definition below) explodes. We relate this observation to a result in Chapter 2 in the remark that follows. The following definition, see Pg. 367 of Bhattacharya and Waymire [13], of the restricted infinitesimal drift and diffusion coefficients does not require the existence of finite moments. Consequently there is no diffusion limit in this case.

Definition: A stochastic process $\{X_s\}$ is a diffusion process if the following holds for every $\gamma > 0$

$$E\Big((X_{s+t} - X_s)I(|X_{s+t} - X_s| \leq \gamma)|X_s = x\Big) = t\mu(x) + o(t); \tag{5.6.75}$$

$$E\Big((X_{s+t} - X_s)^2 I(|X_{s+t} - X_s| \leq \gamma)|X_s = x\Big) = t\sigma^2(x) + o(t); \tag{5.6.76}$$

$$P\Big(|X_{s+t} - X_s| > \gamma|X_s = x\Big) = o(t). \tag{5.6.77}$$

THEOREM 14. If the proposal follows a standard Cauchy distribution then the (restricted) infinitesimal drift and diffusion coefficients do not exist.

Proof: Following the earlier notation let us define for any fixed $\gamma > 0$,

$$b_{n,1}(\mathbf{y},t) \;=\; nE\Big(X_n(\frac{i+1}{n}) - X_n(\frac{i}{n})I(|X_n(\frac{i+1}{n}) - X_n(\frac{i}{n})| \leq \gamma)|\mathbf{Y}_n(\frac{i}{n}) = \mathbf{y}\Big)$$

Given the value of $X_n(\frac{i}{n}) = x$ and $\theta_n(\frac{i}{n}) = \theta$, the set $B_{n,\gamma} = \{|X_n(\frac{i+1}{n}) - X_n(\frac{i}{n})| < \gamma\} = \{\epsilon : |\epsilon| < \frac{\sqrt{n}\gamma}{\theta}\}$.
Therefore, decomposing the expectation over the set A_n and A_n^c as in Section 3.5 we have that

$$\begin{aligned}
b_{n,1}(\mathbf{y},t) &= \sqrt{n}\theta \int\limits_{A_n\cap B_{n,\gamma}} \epsilon f(\epsilon)d\epsilon \\
&+ \sqrt{n}\theta \int\limits_{A_n^c\cap B_{n,\gamma}} \frac{\psi(x+\frac{1}{\sqrt{n}}\theta\epsilon)}{\psi(x)}\epsilon f(\epsilon)d\epsilon, \\
&\quad \text{where } f(\cdot) \text{ is the density of Cauchy(0,1)}, \\
&= \sqrt{n}\theta \int\limits_{A_n\cap B_{n,\gamma}} \epsilon f(\epsilon)d\epsilon + \sqrt{n}\theta\Big(\int\limits_{A_n^c\cap B_{n,\gamma}} \epsilon f(\epsilon)d\epsilon \\
&+ \frac{\theta}{\sqrt{n}}\frac{\psi'(x)}{\psi(x)} \int\limits_{A_n^c\cap B_{n,\gamma}} \epsilon^2 f(\epsilon) + O(\frac{1}{n})\Big), \text{ by Taylor's expansion}, \\
&= \sqrt{n}\theta \int\limits_{B_{n,\gamma}} \epsilon f(\epsilon)d\epsilon + \theta^2\frac{\psi'(x)}{\psi(x)}\Big(\int\limits_{A_n^c\cap B_{n,\gamma}} \epsilon^2 f(\epsilon)d\epsilon + O(\frac{1}{\sqrt{n}})\Big) \\
&= \theta^2\frac{\psi'(x)}{\psi(x)}\Big(\int\limits_{A_n^c\cap B_{n,\gamma}} \epsilon^2 f(\epsilon)d\epsilon + O(\frac{1}{\sqrt{n}})\Big),
\end{aligned}$$

since the standard Cauchy is symmetric about 0. Now if $x < 0$

$$\lim_{n\to\infty} b_{n,1}(\mathbf{y},t) \;=\; \theta^2\frac{\psi'(x)}{\psi(x)}\int_0^\infty \epsilon^2 f(\epsilon)d\epsilon = \infty,$$

If $x > 0$ then the range of the last integral is $(-\infty, 0)$ and its value is also ∞ since the second moment of standard Cauchy is infinity. Therefore the infinitesimal drift coefficient does not exist. Similar things can be shown for the infinitesimal diffusion coefficient. This proves the Theorem. ■

REMARK 45. This behaviour of the Cauchy random variable as the proposal has a heuristic explanation with a result that we have proved in Chapter 2. In that chapter we looked at the Self-Normalized sums (SNS) defined as $\frac{X_1+X_2+\cdots X_n}{\sqrt{X_1^2+X_2^2+\cdots X_n^2}}$. The logic behind defining in this fashion is that even if X_i possess moments which are not finite, the SNS will always have finite variance. This follows from the normalization of the sum of the $X_i^{\prime}s$. Therefore, in a sense, the SNS are the most conservative normalization of the sums that can be expected from any random or non-random normalization. In Chapter 2 we looked at the process version of the above sum and show that the same is true only when the X_i comes from the domain of attraction of a Stable distribution with $\alpha = 2$ (i.e from DAN), see Theorem 1 of Chapter 2. Now if one looks at the diffusion approximation as in Equation (3.3.6), then it is clear that the value of the process X is nothing but the cumulative addition of random variables which has the proposal distribution as its distribution which is normalized by another parameter θ, as given in (3.3.7). In the case of Cauchy distribution $\alpha = 1$. We have proved in Section 2.4.3 of Chapter 2 that if $\alpha < p$ (where in the present case $p = 2$) then the process is not tight. This roughly corresponds to the fact that if one looks at very small sub-interval of the process then the probability that the maximum difference of the process in any sub-interval is greater than any small quantity cannot be small. This is manifested by the fact that the infinitesimal expected deviation and variance, even in the restricted form, as defined in Equations (5.6.75) and (5.6.76) become unbounded as $n \to \infty$.

5.7 Summary

This chapter deals with general choices of target and proposal densities for which the diffusion approximation procedure can be applied. In Section 5.2.1 we assume that the target distribution satisfies the growth condition given in Equation (5.2.1). Under this assumption we consider three possible scenarios for the target which satisfies Equations (5.2.2) to Equation (5.2.4). Under the first condition the m.g.f exists for all $t \in \mathbb{R}$. We show that that under this condition the m.g.f of the limiting distribution of the SDE defined in Equation (5.2.15) also exists for all $t \in \mathbb{R}$. For the second condition the m.g.f exists in a neighbourhood of zero. We show that the corresponding solution has a m.g.f that exists in a neighbourhood of zero. For condition (5.2.4) the m.g.f does not exist in any neighbourhood of zero. This condition is satisfied by the heavy tailed target distribution like the Cauchy and the Pareto type which we deal in Section 5.2.4.

We show that the second moment of X_t is uniformly bounded in t. Thus the limiting distribution, if it exists, cannot be standard Cauchy. This holds because we start with an initial distribution which has bounded second moments. However, we do not have a clear understanding what happens to the limiting distribution of X_t if the target density $\psi(\cdot)$ satisfy (5.2.3) with $\alpha \in (0, 1)$. However, we feel that the limiting distribution is $\psi(\cdot)$ only when its m.g.f exists at least in a neighbourhood of zero. This result is yet to be proved. We hope to study this in the future.

Next we consider what can be the proper choice of the proposal distribution for the Random Walk Metropolis Hastings algorithm. We show that any distribution with the whole real line as the support, symmetric and having finite second moment is a suitable candidate for proposal . We obtain the diffusion approximation which resembles the SDE in Equation (3.4.1) but the constants are different. We also show why a Heavy tailed distribution like the Cauchy is not a suitable candidate for a proposal even if we consider the (restricted) instantaneous drift and diffusion coefficients. We connect this observation with a result that we derived in Chapter 2. We also obtain the diffusion approximation for the multivariate normal proposal and a general multivariate target density.

Chapter 6

Concluding remarks and Future directions

This thesis deals with some recursive equations that arises in theoretical and applied probability. The AMCMC that was defined in Chapter 1 is the sum of random variables from the proposal distribution (normalized by a scaling factor θ_n, which again is a random variable that changes with each iteration). We also observe from Theorem 14 of Chapter 5 that the Cauchy distribution cannot be chosen as the proposal distribution. This leads us to investigate what are the random variables which when normalized by itself (actually by an estimate of the variance computed from the sample) can have a non-trivial distribution. Such objects are called Self-Normalized Sums (SNS). The SNS are the most conservative when it comes to the existence of all finite moments for fixed sample size. In fact the second moment of the SNS is finite even if the random variable themselves do not have any finite moment.

In Chapter 2 we construct a functional form of the SNS called Self-Normalized Process (SNP) and prove weak convergence of the process. We show in Theorem 1 of the same chapter that the only case when we can have a non-trivial distribution is when the random variables are from the domain of attraction of the Normal family and the order of the normalization is two. Since the Cauchy random variables belong to a Stable family with $\alpha = 1$ we do not get convergence to any non-trivial distribution with any order of normalization. This corroborates with the findings in Chapter 5 related to the fact that diffusion approximation cannot be done with the Standard Cauchy as the proposal distribution (see Remark 45). The content of the second chapter is taken from Basak and Biswas [7].

We can view the Adaptive MCMC, X_n as a recursive equation where the increments follow a certain distribution called the proposal distribution. The increments are added with a probability, called the acceptance probability, which again depends on the proposal and the target distribution. This special choice of the acceptance probability ensures that X_n is a Markov chain and has the limiting distribution same as the target distribution.

Sometimes to facilitate faster convergence, the increments are scaled by a factor that changes with each iteration. A related enquiry is whether under this changed scenario the convergence also holds true. Verifying that the convergence holds requires checking some sufficient conditions due to Roberts and Rosenthal [57], see Proposition 3 in Chapter 1. However, verifying those sufficient conditions is not straightforward in all scenarios.

In Chapter 3 we apply the diffusion approximation mechanism to the discrete process. This is actually a scaled down version of the discrete time process. We then focus on the continuous version of the chain. In some of the specific cases the limiting diffusion are investigated. One such case is dealt in Chapter 4 where the choice of the proposal and the target distribution is Normal (0,1).

In Chapter 5 we consider general choices of the target and the proposal distribution for which we can get a diffusion approximation. We consider three classes of density functions that satisfies the growth condition of the solutions of the SDE. These classes correspond to the rate of decay of $x\frac{\psi'(x)}{\psi(x)}$. The rate of decay characterises the existence or the non existence of the m.g.f of the distribution. We show that the condition which ensures the existence of the m.g.f of the target density for all $t \in \mathbb{R}$ also ensures the existence of the m.g.f of the invariant distribution of the SDE for all $t \in \mathbb{R}$. The condition which ensures the existence of the m.g.f of $\psi(\cdot)$ in a neighbourhood of 0 also ensures the same for the m.g.f of the corresponding invariant density. We feel that future research should be able to identify the limiting distribution of the SDE as $\psi(\cdot)$ in such cases. Since the conditions are not true for distributions with heavy tails, such distributions cannot be recovered as the limiting distribution of the diffusion.

We further show that choosing a heavy tailed distribution as the proposal is not appropriate to generate samples from the same (heavy tailed) distribution. The diffusion approximation procedure is also carried out in detail for multivariate target densities. In the case of standard MCMC we can apply the Stein's lemma to show that the limiting

distribution of the corresponding diffusion is the same as the target distribution $\psi(\cdot)$.
Some tasks awaits future research. In Chapter 2 the rate of convergence of the SNP should prove to be important. A non-uniform Berry Essen bound was given in Bentkus and Gotzë [10], when the random variables are from DAN, and a bound using Saddlepoint approximation was proved in Jing *et al.* [36]. Although the process convergence is for $p = \alpha = 2$, Logan *et al.* [40] have shown that the Self-Normalized sums can converge for $p > \alpha \in (0, 2)$. Using their techniques we have shown in Chapter 2 what the possible limiting characteristic distribution of the finite dimensional distribution of the SNP would look like. From our personal communication with Qi-Man Shao we came to know about an unpublished result on limiting finite-dimensional distribution of $((S_{[nt_1]}/V_{n,p}, \ldots, S_{[nt_k]}/V_{n,p}), \ \ p > \alpha)$ where they have shown that the limiting joint distribution is a mixture of Poisson-type distribution using technique of Csörgő and Horvath [18]. The rate of convergence for this case has not been explored to our knowledge. In Chapter 4 we have shown explicit convergence in the case of the Normal target densities. However it will be important to know when light tailed distributions other than the Normal distribution is appropriate as the target distribution. In this context, Stein's characterization of distribution other the Normal may be useful to find the limiting distribution of X_t. Although the diffusion approximations are obtained in the thesis, finding the explicit convergence to the target density would be ideal.
In Chapter 5 we have been able to characterize the situation where this technique of diffusion approximation cannot be applied. However, for special choices of the target distribution further research is needed to identify the limiting distribution with the target density. This may not be possible using the method of moments since computing the moments for general target densities is not straightforward. Further, if we want to apply the Stein's method to the AMCMC then one needs to compute the exact joint expectation of $f'(X_s)\theta_s$ for some suitable class of function f. We plan to do this this in our future research.

Further, Normal proposal in case of a Heavy tailed distribution ψ or multi-modal densities is not always very efficient. Algorithms must be designed to handle such cases. It should be investigated whether combining the ideas of adaptability and heavy tailed proposal increases the efficiency. There are some work in this area by Jarner *et al.* [35]. For theoretical comparison between two or more adaptive chains one needs to obtain computable bounds on the time taken to reach stationarity; see Joulin and Ollivier [37] for bounding times for Standard MCMC. We have some ideas on these for the Normal target density (see Remark 24 of Chapter 4) which may be extended to other general admissible target densities.

Bibliography

[1] Andrieu C. and Moulines E. (2006). On the ergodic properties of some Adaptive MCMC algorithms. *The Annals of Applied Probability* **16, 3**, 1462-1505.

[2] Athreya K. B. and Lahiri S. (2006). Measure Theory and Probability Theory. *Springer.*

[3] Barbour A. D. (1990). Stein's method for Diffusion Approximation. *Probab. Th. Rel. Fields.* **84**, 297-322.

[4] Basak G. K. and Dasgupta A. (2013). An Ornstein Uhlenbeck process associated to Self-Normalized sums. arXiv: 1302.0158v1.

[5] Basak G. K. and Biswas A. (2013). Diffusive limits for Adaptive MCMC for Normal target densities. arXiv: 1301.4030.

[6] Basak G. K. and Biswas A. (2012). Comparison between Standard and Adaptive MCMC via their diffusion limits. Link at http://arxiv.org/abs/1201.1433.

[7] Basak G. K. and Biswas A. (2013). Process convergence of self-normalized sums of i.i.d. random variables coming from domain of attraction of stable distributions. *Proc. Indian Acad. Sci. (Math Sci.)* **123, 1**, 85 - 100.

[8] Bedard M. and Rosenthal J. R. (2008). Optimal Acceptance Rates for Metropolis-Hastings Algorithms: Heading towards general target distributions. *Canad. J. Statist.* **36**, 483-503.

[9] Bedard M. (2006). On the robustness of optimal scaling in Metropolis Hastings algorithm, Ph. D dissertation, University of Toronto.

[10] Bentkus V. and Götze, F. (1996). The Berry-Esseen bound for Student's statistics, *Ann. Probab.* **24**, 491 - 503.

[11] Bhattacharya R. N. (1978). Criteria for Recurrence and Existence of Invariant measures for Multidimensional Diffusions, *Ann. Probab.*, **6,4**, 541 - 553.

[12] Bhattacharya R. N. and Ramasubramanian S. (1982). Recurrence and Ergodicity of Diffusions, *Journal of Multivariate Analysis*, **12**, 95 - 122.

[13] Bhattacharya R. N. and Waymire E. (2009). Stochastic Processes with Applications, SIAM.

[14] Billingsley P. (1999). Convergence of Probability Measures, Wiley series in Probability and Statistics.

[15] Brooks S., Gelman A., Jones G. and Meng X. L. (eds) (2011). Handbook of MCMC, CRC Press.

[16] Chatterjee S.(2014). A short survey of Stein's method. *arXiv:1404.1392v1[math.Pr].*

[17] Csörgő M., Szyszkowicz B. and Wang Q. (2003). Donsker's Theorem for Self-Normalized partial sums processes. *Ann. Probab.* **31, 3**, 1228-1240.

[18] Csörgő M., Horvath L. (1988). Asymptotic representation of Self-Normalized sums,*Probability and Mathematical Statistics*, **9.1**, 15-24.

[19] Darling D. A. (1952). The influence of the maximum term in the addition of independent Random Variables, *Transactions of the American Mathematical Society* **73**, 1, 95 - 107.

[20] Efron B. (1969). Student's t-test under symmetry conditions. *J. Am. Statist. Assoc.* **64**, 1278- 1302.

[21] Embrechts P., Klüppelberg C., Mikosch T. (1997). Modelling Extremal Events for Insurance and Finance, Springer-Verlag.

[22] Ethier S. N. and Kurtz T. G. (2000). Markov Processes: Characterization and Convergence, Wiley Series in Probability and Statistics.

[23] Feller W. (1971). An introduction to Probability Theory and its Applications, Vol 2, 2nd edition, Wiley, New York.

[24] Feller W. (1952). The parabolic differential equations and the associated semi-groups of transformations, *Annals of Mathematics*, **55,3**, 468-519.

[25] Gelman A., Gilks W. R. and Roberts G. O. (1997). Weak convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *Annals of Applied Probability* **7, 1**, 110-120.

[26] Gelman A. and Pasarica C. (2010). Adaptively scaling the Metropolis Hastings algorithm using the Expected Square Jumping Distance. *Statistica Sinica* **20, 1**, 110 - 120.

[27] Giné E., Götze F. and Mason D. M . (1997). When is the Students' t asymptotically standard Normal? *Ann. Probab.* **25**, **3**, 1514-1531.

[28] Gosset (Student) W. S. (1908). On the probable error of a mean . *Biometrika* **6**, 1-25.

[29] Griffin P. S. and Kuelbs J. D. (1989). Self-Normalized law of the iterated logarithm. *Ann. Probab.* **17,4**, 1571 - 1601.

[30] Griffin P.S. and Kuelbs J. D. (1991). Some extension of the laws of iterated logarithms via Self-Normalization. *Ann. Probab.* **19, 1**, 380 - 395.

[31] Griffin, P. S. (2002). Tightness of the Student t statistic. *Elect. Comm. Probab.* **7**, 181 - 190.

[32] Hastings W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57(1)**, 97-109.

[33] Haario H., Saksman E. and Tamminen J. (2001). An adaptive Metropolis algorithm, *Bernoulli* **7(2)**, 223-242.

[34] Hörmander Lars (1967). Hypoelliptic second order differential equations, *Acta Math.* **119**, 147-171.

[35] Jarner S. F. and Roberts G. O. (2007). Convergence of Heavy-tailed Monte Carlo Markov Chain Algorithms. *Scand. J of Statistics* **34, 4**, 781-815.

[36] Jing B. Y., Shao Qi-Man, Wang Q. (2004). Saddlepoint approximation for Students t statistics with no moment condition. *The Annals of Statistics* **32,6**, 2679-2711.

[37] Joulin A. and Ollivier Y. (2010). Curvature, Concentration and Error estimates for Markov chain Monte Carlo. *The Annals of Probability* **38, 6**, 2418-2442.

[38] Kliemann W. (1987). Recurrence and Invariant Measures for Degenerate Diffusions. *The Annals of Probability* **15, 2** , 690-707.

[39] Kloeden P. E. and Platen E. (1992). Numerical Solution of Stochastic Differential Equations, Springer.

[40] Logan B.F., Mallows C. L., Rice S. O. and Shepp, L. A. (1973). Limit distribution of Self-Normalized sums. *Ann. Probab.* **1**, 788 -809.

[41] Mao X. (1999). Stability of stochastic differential equations with Markovian switching. *Stochastic Processes and their applications* **79,1**, 45-67.

[42] Marshall T. and Roberts G. (2012). An adaptive version to Langevin MCMC. *Stat. Computing* **22**, 1041-1057.

[43] Mattingly J. C., Pillai N. and Stuart A. M. (2010). Diffusive Limits of Random Walk Metropolis Algorithm in High Dimension, arXiv: 1003.43006v3.

[44] Metropolis N., Rosenbluth A., Rosenbluth M., Teller A. and Teller E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087-1091.

[45] Meyn S. P. and Tweedie R. L. (1993). *Markov Chains and Stochastic Stability.* Springer. London.

[46] Meyn S. P. amd Tweedie R. L. (1993). Stability of Markovian Processes II: Continuous-time process and sampled chains, *Adv. Appl. Prob.* **25**, 487-517.

[47] Monro S. and Robbins H. (1951). A Stochastic Approximation Algorithm. *The Annals of Mathematical Statistics* **22, 3**, 400-407.

[48] Nelson D. B.(1990). ARCH models as diffusion approximation . *Journal of Econometrics* **45,1-2**, 7-38.

[49] Nualart D. (2006). The Malliavin calculus and related topics, Springer-Verlag.

[50] O'Brien, G. L. (1980). A limit Theorem for sample maxima and heavy branches in Galton Watson trees. *J. Appl. Probab.* **17**, 539 - 545.

[51] Oksendal B., (2000), *Stochastic Differential Equations*, Springer-Verlag.

[52] Peña, Victor H., Lai, Tze Leung and Shao, Qi-Man (2009). Self-Normalized Processes: Limit Theory and Statistical Applications, Springer.

[53] Račkauskas A. and Sequet C. (2001). Invariance principles for adaptive self-Normalized partial sums processes. *Stochastic Process. Appl.* **95**, 6381.

[54] Roberts G. O. and Rosenthal J. S. (2001). Optimal Scaling for various Metropolis-Hastings algorithms. *Statist. Sci.* **16, 4**, 351-367.

[55] Roberts G. O. and Rosenthal J. S. (2004). General State Space Markov chains and MCMC algorithms. *Probability Surveys* **1**, 20-71.

[56] Roberts G. O. and Rosenthal J. S. (2006). Examples of Adaptive MCMC, *Journal of Computational and Graphical Statistics* **18**, 2,349-367.

[57] Roberts G. O. and Rosenthal J. S. (2007). Coupling and Ergodicity of Adaptive MCMC. *J. Appl. Probab.* **44, 2**, 458-475.

[58] Roberts G. O. and Tweedie R. L. (1996). Exponential convergence of Langevin diffusions and their discrete approximations, *Bernoulli* **2**, 341-364.

[59] Samorodnitsky G. S. and Taqqu M.(1994). Stable non-Gaussian random process: stochastic models with infinite variance. Chapman and Hall.

[60] Senata E. (1976). Regularly Varying Functions. Lecture Notes in Mathematics, Springer-Verlag.

[61] Shao Qi-Man (1997). Self-Normalized large deviations. *Ann. Probab.* **25, 1**, 285 - 328.

[62] Skorokhod A. V. (1989). Asymptotic Methods in the Theory of Stochastic Differential Equations. American Mathematical Society, Providence, Rhode Island.

[63] Stramer O. and Tweedie R. L. (1999a). Langevin type models I: Diffusion with given stationary distribution and their discretization. *Methodology and Computing in Applied Probability* **1:3**, 283-306.

[64] Stramer O. and Tweedie R. L., (1999b). Langevin type models II: Self targeting candidates for MCMC algorithms. *Methodology and Computing in Applied Probability* **1:3**, 307-328.

[65] Stroock D. W. and Varadhan S. R. S. (1979). Multidimensional Diffusion Processes, Springer-Verlag, Berlin.