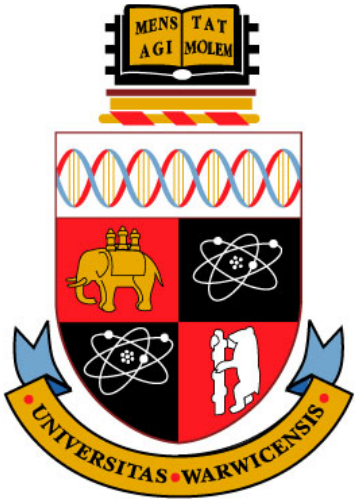


Exact Algorithms for simulation of diffusions with discontinuous drift and robust Curvature Metropolis-adjusted Langevin algorithms

by

Katarzyna Barbara Taylor

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

June 2015

THE UNIVERSITY OF WARWICK

Contents

Acknowledgments

I would like to thank my PhD supervisor Gareth Roberts for inspirational discussions. I am grateful for an opportunity to see statistics through his experience, broad knowledge and innovative thinking. Also for his support and encouragement, especially during difficult times. Many thanks to Krzysztof Łatuszyński for working with me on CMALA project. I highly appreciate his good advice and sharing his insight. I am grateful to Omiros Papaspiliopoulos for his contagious enthusiasm, sharing his expertise while working together on Exact Algorithms and invitation for research visit in Barcelona. I would like to thank Andrew Stuart for his support and understanding during my last year of PhD studies.

I feel very privileged that I could participate in the great scientific life of the Department of Statistics at Warwick and also Warwick Mathematics Institute. The Department of Statistics has been very generous to support my participation in conferences and workshops and I acknowledge the Centre for Research in Statistical Methodology (CRiSM) for financially supporting my research.

I would like to thank my Mum for her unconditional love and support:
Dziękuję Mamo za poświęcony czas, uwagę i cierpliwość, dziękuję za pomoc w realizacji marzeń!

My last but not least thanks go to friends and colleagues who shared fun with me during my PhD studies and especially to Henry and Peter for being great friends during the journey.

Declarations

In this thesis we have used content from two papers in preparation:

"Robust Curvature MALA algorithms" K. Łatuszyński, G.O. Roberts and K. B. Taylor

"Exact Algorithms for simulation of diffusions with discontinuous drift", O. Papaspiliopoulos, G.O. Roberts and K. B. Taylor.

Abstract

In our work we propose new Exact Algorithms for simulation of diffusions with discontinuous drift and new methodology for simulating Brownian motion jointly with its local time. In the second part of the thesis we introduce Metropolis-adjusted Langevin algorithm which uses local geometry and we prove geometric ergodicity in case of benchmark distributions with light tails.

Chapter 1

Introduction

The thesis is a result of my work on two independent projects. Both of them aim at formulating efficient algorithms for sampling from probability distributions. However the distributions in each case are defined on significantly different spaces and the sampling requires applying conceptually distinct techniques.

In the first project we focus on simulation of diffusions. To avoid confusion let us explain straight away that by diffusion we mean solution to a specific class of stochastic differential equations. Diffusions that we concentrate on are stochastic processes with random variables indexed by t in some interval $[0, T]$ and t is often referred to as *time*. The realisations of the process are continuous functions on $[0, T]$ referred to as *paths*. Here we are interested in producing paths according to diffusion measure and/or sampling from finite-dimensional distributions of diffusions.

The second project is about sampling from low-dimensional distributions whose support is usually $\mathbb{R}^d$ with densities known up to a multiplicative constant. We focus on Metropolis-adjusted Langevin algorithms which use geometry of the density to construct proposals.

We decided to discuss the projects as two separate parts of the thesis because the character of simulation and methodology applied are quite different. In the first part we deal with infinite dimensional objects and the algorithms allow for **exact** simulation, i.e. involving only computer representation errors but not sampling from approximate distributions. While the other one aims at sampling from well chosen 'similar' distributions to the *target* one. This method produces results which quality depends on computational resources allocated. While the other one from the beginning produces exact results and more computational resources allow to produce realisations on longer time intervals and allow to refine the obtained structure.The main ideas used to address mentioned simulation problems are different. Namely in one of them the geometry of density function is used to find good proposals for Metroplis-Hastings algorithm which produces a Markov chain whose limiting distribution is our *target* one. While in the other problem smartly constructed auxiliary random variables together with special bridge simulations involving local time process allow to perform rejection sampling which would not otherwise be possible due to the necessity to use information about whole diffusion paths. The last but not the least important difference is that in one simulation we get independent realisations while in the other not. We discuss both cases in detail in the respective following parts of the thesis. We would like to stress that our interest lies in the theoretic analysis of simulation methodology.

Our motivation to do research on simulation methods for diffusions is two-fold. Nowadays diffusions are a basic tool to model random processes developing in time. They have been proved useful in a very broad spectrum of disciplines, starting from physics and finance through engineering, meteorology, geology, neuroscience, microbiology and many other. Diffusions satisfying the following SDE

$$dX_t = \alpha(X_t)dt + dB_t, \quad t \in [0,T], X_t = x$$

with discontinuous drift function α depending on X_t may be used to describe

processes whose dynamics change radically when X_t hits certain levels. This can be encountered for example in geology when we model flow through different media. Diffusions with discontinuous drift arise also as limiting distributions in queueing theory. Possibility to sample diffusions is important in construction of MLE estimators and in more general context of inference for diffusions.

There has been high demand in applied science community for efficient methods for simulating diffusions for decades. However most known methods use variety of discretisation/approximation schemes, like Euler or Shoyi-Ozaki schemes, survey of methods can be found e.g. in [KP92]. The quality of iterative traditional schemes deteriorates when simulation needs to be carried for long time intervals. The approximation errors are usually amplified with time. Our second motivation is the very innovative idea behind Exact Algorithms which is interesting in its own right.

Interest in the same problem as ours can be found in the following articles [ÉM13b], [EM13a]. However their approach to solving the problem is different than suggested in the thesis.

We would like to list here our main contributions to addressing aforementioned simulation problems

- construction of three algorithms: Alg.1, Alg.2 and Alg.3 will allow for **exact** simulation of diffusions with discontinuous drift and ensure the quality of simulation does not deteriorate with the increase of the length of the interval $[0, T]$
- the algorithms allow for simulation of both conditional and unconditional diffusions where, by conditional, we mean the ones whose paths are hitting

prescribed value at prescribed time

- Alg.3 addresses the simulation problem in the harder case of 'negative jump' at discontinuity in the drift function
- the structure of the sampled paths can be refined 'offline' which is important for inference for diffusions
- methodology for simulation of various bridges involving Brownian motion and its local time
- we derive MALA algorithms with state-dependent scaling
- we prove geometric ergodicity for benchmark light-tailed distributions.

.

Part I

Exact Algorithms for simulation of diffusions with discontinuous drift

Chapter 2

Preliminaries

2.1 Introduction to Exact Algorithms

As we have already mentioned in Chapter 1, the problem of simulation of diffusions has been of great interest to many sectors of industry and business where data modelling is necessary, like finance for example. We are especially interested in Exact Algorithms because of their innovative approach to simulation of solutions of SDEs. Classical methods for simulation of diffusions include approximation errors and quality of simulation deteriorates over long time intervals which is not the case for Exact Algorithms.

The breakthrough article where the first Exact Algorithm was introduced was [BR05]. However it has since been improved and we discuss here the version of EA1 as in [BPR06]. In [BPR08] a general framework for Exact Algorithms can be found. We concentrate here on EA1 though. Let us present now the set of assumptions for the algorithm EA1.

Set of conditions for EA1

Assumption 1. *There exists nonexplosive weak solution to SDE in (2.1)*

Assumption 2. *The Girsanov theorem may be applied and the local martingale is a true martingale.*

Assumption 3. α *is a differentiable function.*

Assumption 4. *Function* $\exp\{A(u) - \frac{(u-x)^2}{2T}\}$ *is integrable.*

Assumption 5. *Function* $(\alpha^2(u) + \alpha'(u))$ *is bounded below*

Exact Algorithms are applicable to a limited class of multi-dimensional diffusions but to a very wide class of one-dimensional processes. The restrictive Assumption 5 has been lifted from EA3, partially already from EA2. However Assumption 3 is still in effect. Our work concentrates on removing this Assumption so that the new algorithm will be applicable to diffusions with piecewise continuous drift functions.

As our developments in simulation of diffusions apply only to one-dimensional diffusions we assume from now on that described diffusions are one-dimensional processes. Let us discuss construction of Exact Algorithm 1 here. The aim is to simulate trajectories of the diffusion X which is a solution of

$$dX_t = \alpha(X_t)dt + dB_t, \quad t \in [0, T], X_t = x \tag{2.1}$$

where B is Brownian motion defined on the underlying probability space $(\Omega, \mathcal{F}, P)$. Diffusions with non-unitary diffusion coefficient need to be first transformed to 2.1 by Lamperti transformation.

Remark 1. SDEs under discussion admit unique strong solutions as it is sufficient that the drift function is bounded and Borel measurable while diffusion coefficient is unitary. It follows from Theorem 1 in [Ver81]

First we introduce the notation and describe the underlying framework for EA1. We denote the space of one-dimensional continuous functions on $[0,T]$ by $C := C([0,T],\mathbb{R})$ and refer to its elements as trajectories or paths. We set $\Omega = C$ and define $X_t(\omega) \equiv \omega(t)$ for $\omega \in \Omega$ and $t \in [0,T]$. X is used for both a coordinate mapping and diffusion but we believe it is clear from the context which one is under discussion. The σ-algebra on C is defined as the cylinder σ-algebra $\mathcal{C} := \sigma(X_t; t \in [0,T])$. Next we set $\mathcal{F}_t^0 \equiv \sigma(\{X_s; s \leq t\})$ for $t \in [0,T]$. $(B_t^x, t \in [0,T])$ is Brownian motion started at x and $\mathbb{W}$ the law of B^x. Further let $\mathcal{F}_t$ be $\mathbb{W}$-augmentation of $\mathcal{F}_t^0$. We denote by $\mathbb{Q}$ the law of diffusion X on $(C,\mathcal{C})$.

Next having performing rejection sampling in mind we find the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{W}$ using Girsanov transformation.

$$\frac{d\mathbb{Q}}{d\mathbb{W}} = \exp\Big\{\int_0^T \alpha(X_t)dX_t - \frac{1}{2}\int_0^T \alpha^2(X_t)dt\Big\}. \tag{2.2}$$

Then by introducing an auxiliary function $A : \mathbb{R} \to \mathbb{R}$, $A(u) := \int_0^u \alpha(y)dy$ and applying Itô's lemma to the Brownian motion starting at x and function A we find that

$$\frac{d\mathbb{Q}}{d\mathbb{W}} = \exp\Big\{A(X_T) - A(x) - \frac{1}{2}\int_0^T (\alpha^2(X_t) + \alpha'(X_t))dt\Big\}$$

Note that Assumption 3 is needed to apply Itô's lemma. The next step in EA1 is to consider biased Brownian motion $\hat{B}$ starting at x and satisfying $\hat{B}_T \sim h$ where $h(u) \propto \exp\{A(u) - \frac{(u-x)^2}{2T}\}$. Note that this requires Assumption 4 to be satisfied. Measure induced on $(C,\mathcal{C})$ by $\hat{W}$ is denoted by $\mathbb{Z}$. We now recall the Proposition 1 from [BPR06].

Proposition 1. *Let $M = \{M_t; 0 \leq t \leq T\}$, $N = \{N_t; 0 \leq t \leq T\}$ be two stochastic processes on $(C,\mathcal{C})$ with corresponding probability measures $\mathbb{M},\mathbb{N}$. Assume that f_M, f_N are the densities of the ending points M_T and N_T respectively with identical support $\mathbb{R}$. If it is true that $(M|M_T = \rho) \stackrel{d}{=} (N|N_T = \rho)$, for all*

$\rho \in \mathbb{R}$, *then:*

$$\frac{d\mathbb{M}}{d\mathbb{N}} = \frac{f_M}{f_N}(X_T). \tag{2.3}$$

Hence

$$\frac{d\mathbb{W}}{d\mathbb{Z}} = \frac{\varphi_{x,T}(X_T)}{h(X_T)} \tag{2.4}$$

where $\varphi_{x,T}(u)$ is density of normal distribution with mean x and variance T. It follows that

$$\frac{d\mathbb{Q}}{d\mathbb{Z}} = \frac{d\mathbb{Q}}{d\mathbb{W}}\frac{d\mathbb{W}}{d\mathbb{Z}} \propto \exp\Big\{ -\frac{1}{2}\int_0^T (\alpha^2(X_t) + \alpha'(X_t))dt \Big\}.$$

Under Assumption 5 we can find $k = \inf_{u\in\mathbb{R}} \frac{\alpha^2(u)+\alpha'(u)}{2}$ and define

$$\phi(u) := \frac{\alpha^2(u) + \alpha'(u)}{2} - k,\ u \in \mathbb{R}.$$

Thus

$$\frac{d\mathbb{Q}}{d\mathbb{Z}} \propto \exp\Big\{ -\int_0^T \phi(X_t)dt \Big\} \leq 1; \quad \mathbb{Z}-a.s. \tag{2.5}$$

The next step involves rejection sampling using $d\mathbb{Q}/d\mathbb{Z}$. We note from (2.5) that the Radon-Nikodym derivative is bounded and it is achieved by choosing biased Brownian motion instead of Brownian motion as a candidate process. However, it is not possible to perform rejection sampling if it involves revealing the whole path. Theorem 1 from [BPR06] allows to circumvent this difficulty. We state a simpler version of this Theorem in Lemma 2.1.

Lemma 2.1. *If ϕ is bounded above by M and Ψ is a marked Poisson process with intensity 1 on $[0,T]\times[0,M]$ and N is the number of points of Ψ found in the hypograph of $\phi(\omega)$, then*

$$P(N = 0|\omega) = \exp\Big\{ -\int_0^T \phi(X_t)dt \Big\}.$$

We note that by the above Lemma it holds that

$$\frac{d\mathbb{Q}}{d\mathbb{Z}} \propto P(N = 0|\omega)$$

and it is sufficient to reveal ω at finite number of times to decide whether the marks lie below or above $\phi(X)$. In the case when all of them lie above the so-called 'skeleton' is accepted. By the skeleton we mean a sample from the finite dimensional distribution of Brownian bridge together with the initial and end points of the bridge. We denote the realization of Ψ by $((\tau_1, \psi_1), \cdots, (\tau_k, \psi_k))$ with $k \in \mathbf{Z}^+$.

If we are interested in the realization of diffusion at some other time instances in $[0, T]$ than $(0, \tau_1, \cdots, \tau_k, T)$ then they can be obtained by simulating Brownian bridges at the times of interest. Note that the path has already been accepted and is generated according to $\mathbb{Z}$. Thus, if we want to reveal more values of X later on, we need to produce a sample according to Wiener measure $\mathbb{W}$ conditioned on the skeleton. To sum up EA1 has the following structure

Exact Algorithm 1 (EA1)

INPUT:

$\alpha(u) : \mathbb{R} \to \mathbb{R}$ drift function

$\{0 = t^0, t^1, \cdots, t^n, t^{n+1} = T\}$ time instances of interest

$X_0 = x$ initial point

I 1. Define

(a) $A : \mathbb{R} \to \mathbb{R}$, $A(u) := \int_0^u \alpha(y)dy$

(b) $k = \inf_{u\in\mathbb{R}} \frac{\alpha^2(u)+\alpha'(u)}{2}$

(c) $\phi : \mathbb{R} \to \mathbb{R}$, $\phi(u) := \frac{\alpha^2(u)+\alpha'(u)}{2} - k$

2. Find the upper bound M of ϕ.

II Generate realisation of X at T and at random times in $(0, T)$:

1. Sample $X_T \sim h(u) \propto \exp\{A(u) - \frac{(u-x)^2}{2T}\}$

2. Sample an auxiliary planar Poisson process Ψ

(i) Sample $k \sim$ Poiss(TM)

(ii) Sample $U_1, \cdots, U_k$ i.i.d., $U_1 \sim$ Uni$((0, T) \times (0, M))$

(iii) Sort $\{u_1, \cdots, u_k\}$ in increasing order of their first coordinate τ to get $((\tau_1, \psi_1), \cdots, (\tau_k, \psi_k))$

3. Use Ψ to perform rejection sampling:

(a) Sample $(X_{\tau_1}, \cdots, X_{\tau_k})$ from the Brownian bridge measure $\mathbb{W}_{0,T}^{x,x_T}$

(b) Compute $\phi(X_{\tau_i})$ for $i \in \{1, \cdots, k\}$

(c) If $\phi(X_{\tau_i}) < \psi_{\tau_i}$ for each $i \in \{1, \cdots, k\}$ then proceed to **III**. Otherwise return to **II**.**1**

III Generate realisation of X at $(t^1, \cdots, t^n)$ conditioned on values of X at $(0, \tau_1, \cdots, \tau_k, T)$:

(a) Sort in increasing order the set $\{0, \tau_1, \cdots, \tau_k, t^1, \cdots, t^n, T\}$ to get $(0, t_1, \cdots, t_{k+n}, T)$

(b) Repeat n times

(i) Pick $i \in \{1, \cdots, k+n\}$ such that X_{t_i} has not been sampled yet

(ii) Find $t_L := \max_{\{0,t_1,\cdots,t_{k+n},T\}}\{t_j$ such that X_{t_j} has been already revealed and $t_j < t_i\}$

(iii) Find $t_U := \min_{\{0,t_1,\cdots,t_{k+n},T\}}\{t_j$ such that X_{t_j} has been already revealed and $t_j > t_i\}$

(iv) Sample X_{t_i} from the Brownian bridge measure $\mathbb{W}_{t_L,t_U}^{x_L,x_U}$

OUTPUT: $(x, x_{t_1}, \cdots, x_{t_{k+n}}, x_T)$ and $(0, t_1, \cdots, t_{k+n}, T)$

2.2 Notes on local time

Informally speaking one-dimensional diffusions can be looked at from two angles. 'Standard perspective' is to investigate how values of a diffusion are changing over infinitesimal time interval. In fact, diffusions can be characterised by their infinitesimal mean and variance. The second 'local time perspective' is to look at a diffusion through a stochastic process associated with it called *local time*. Here the focus is on the amount of time when the values of the diffusion belong to some infinitesimal interval. It appears that using connection between these two approaches is crucial for simulating diffusions with discontinuous drift.

Here we discuss local times with a view of their application to our algorithms. We answer the question why exactly we need to resort to local time process in Chapter 4. Let $(B_t)_{t\geq 0}$ be Brownian motion with $(\Omega, \mathcal{F}, \mathrm{P})$ the underlying probability space.

There exist several equivalent definitions of Brownian local time. We present three of them as we find them the most useful in the context discussed; for more details see e.g. [RW00], [RY99], [Kar91], [MP10].

Definition 2.2. *Let L be a stochastic process such that for all $t \geq 0$ and $x \in \mathbb{R}$*

$$L(t,x) := \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{(x-\varepsilon, x+\varepsilon)}(B_s) ds$$

where $(B_t)_{t \geq 0}$ is one-dimensional Brownian motion. Then L is a Brownian local time.

Definition 2.3. *Let B be Brownian motion and r a real number. Then an increasing continuous process L^r satisfying*

$$|B_t - r| = |B_0 - r| + \int_0^t sgn(B_s - r) dB_s + L_t^r,$$

where $sgn(x) = 1$ for $x > 0$, $sgn(x) = -1$ for $x \leq 0$, is called the Brownian local time at r.

Definition 2.4. *Let $\{L_t^r\}$ with $t \in \mathbb{R}_{\geq 0}$ and $r \in \mathbb{R}$ be a process with the following properties*

1. *The mapping $(t,r) \mapsto L_t^r$ is jointly continuous*
2. *The **occupation density formula** holds, i.e.*

$$\int_0^t f(B_s) ds = \int_{-\infty}^{\infty} f(r) L_t^r dr$$

for all $t \geq 0$ and bounded measurable functions f

*then the process $(L_t^r)_{t \geq 0,\, r \in \mathbb{R}}$ is called **Brownian local time**.*

Last definition is valid together with Trotter's theorem which states that a stochastic process defined in Definition 2.4 actually exists. There is corresponding to Definition 2.3 definition of local time for continuous semimartingales. In Theorem 2.5 we combine Tanaka formula and result of M. Yor on continuity.

Theorem 2.5. *Let X be a continuous semimartingale. Then there exists for each $r \in \mathbb{R}$ an increasing continuous process L^r satisfying*

$$|X_t - r| = |X_0 - r| + \int_0^t sgn(X_s - r) dX_s + L_t^r,$$

where $sgn(x) = 1$ for $x > 0$, $sgn(x) = -1$ for $x \leq 0$. Furthermore there exists a version of $\{L_t^r : r \in \mathbb{R}, t \in (0, \infty)\}$ such that it is jointly continuous in t and righ-continuous with left limits in r. We call it the local time of X.

Application of the following theorem due to H.F. Trotter and P.A. Meyer is crucial for our simulations

Theorem 2.6. *Let $f : \mathbb{R} \to \mathbb{R}$ be a difference of two convex functions with a second derivative measure f'' and X is a continuous semimartingale. Then we have a.s.*

$$f(X_t) = f(X_0) + \int_0^t \frac{f'_+(X_s) + f_-(X_s)}{2} dX_s + \frac{1}{2} \int_{-\infty}^{\infty} L_t^r f''(dr).$$

Furthermore for any bounded measurable f it holds that

$$\int_0^t f(X_x) d[X]_s = \int_{-\infty}^{\infty} L_t^r f(r) dr.$$

Proofs of the above Theorems can be found e.g. in [RW00]. We followed presentation of theory on local times therein.

We use multiple times in the following Chapter formulas for joint density of Brownian motion and its local time (see e.g. [BS02]):

$$P_x(B_s \in \mathrm{d}b, L_s^r = 0) = \frac{1}{\sqrt{2\pi s}} e^{-\frac{(b-x)^2}{2s}} \mathrm{d}b - \frac{1}{\sqrt{2\pi s}} e^{-\frac{(|b-r|+|r-x|)^2}{2s}} \mathrm{d}b \tag{2.6}$$

and for $l > 0$ we have

$$P_x(B_s \in \mathrm{d}b, L_s^r \in \mathrm{d}l) = \frac{1}{s\sqrt{2\pi s}} (l + |b - r| + |r - x|) e^{-\frac{(l+|b-r|+|r-x|)^2}{2s}} \mathrm{d}b\mathrm{d}l \tag{2.7}$$

We define functions $f_s^{x,r,const} : \mathbb{R} \to \mathbb{R}_+$ and $f_s^{x,r} : \mathbb{R} \times (0,\infty) \to \mathbb{R}_+$ for each $s \in (0,T]$ satisfying

$$P_x(B_s \in \mathrm{d}b, L_s^r = 0) = f_s^{x,r,const}(b)\ \mathrm{d}b$$
$$P_x(B_s \in \mathrm{d}b, L_s^r \in \mathrm{d}l) = f_s^{x,r}(b,l)\ \mathrm{d}b\mathrm{d}l.$$

Both of these functions are probability densities with respect to Lebesgue measure, in one and two dimensions respectively. Similarly we define other probability densities in Chapter 3 including conditional ones. Where needed we use additional intuitive sub- and superscripts.

Remark 1. An important property of Brownian motion and its local time that allows us for various bridges simulations is that Brownian motion jointly with its local time constitute a Markov additive process. (See e.g. [RY99] Chapter X.) It allows us to write that

$$\mathcal{L}(B_t, L_t^r(B)) = \mathcal{L}(B_s, L_s^r(B)) * \mathcal{L}(B_{t-s}, L_{t-s}^r(B)) \quad \text{for } s < t.$$

Chapter 3

Bridge simulations

In the algorithms defined in Chapter 4 we exploit heavily properties of Brownian motion. In particular we use regular conditional probabilities with conditioning on events involving local time. We suggest several sampling methods from distributions which will be crucial for algorithms Alg.1, Alg.2 and Alg.3. We recall a standard result about Brownian bridges.

3.1 Brownian bridge simulation: $B^x \sim \mathbb{BB}^{x,b_1,b_3}_{s_1,s_2,s_3}$

Let $0 \leq s_1 < s_2 < s_3 \leq T$. B^x can be looked at as a random variable on $C([0,T])$ and it applies also to B^x given $\sigma(B^x_{s_1}, B^x_{s_3})$. The process B^x conditioned on $A := \{B^x_{s_1} = b_1, B^x_{s_3} = b_3\}$ is called a *Brownian bridge satisfying* A and we denote its measure on $C([0,T])$ by $\mathbb{BB}^{x,b_1,b_3}_{s_1,s_2,s_3}$.

In this and all following bridge simulations if $s_1 = 0$ then we consider only $b_1 = x$ as all the paths of B^x start at x.

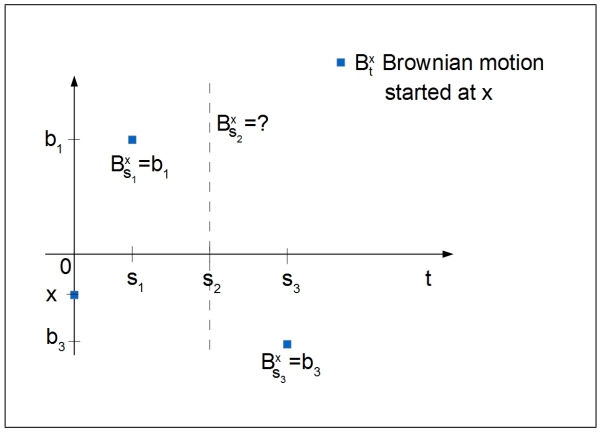


Fig. 1 Unveiled information about Brownian bridge (defining conditions).

This and future diagrams illustrate information included in conditioning and also indicate realisation of which variable(s) we are especially interested in.

Density of $B_{s_2}^x$ given A can be easily obtained by using Bayes' Theorem, the Markov property and time-homogeneity of B

$$
\begin{aligned}
&P(B_{s_2}^x \in \mathrm{d}b_2 | B_{s_1}^x = b_1, B_{s_3}^x = b_3) \\
&= \frac{f_{s_2}^x(b_2|B_{s_1}^x = b_1) f_{s_3}^x(b_3|B_{s_1}^x = b_1, B_{s_2}^x = b_2)}{f_{s_3}^x(b_3|B_{s_1}^x = b_1)} \mathrm{d}b_2 \\
&= \frac{f_{s_2}^x(b_2|B_{s_1}^x = b_1) f_{s_3}^x(b_3|B_{s_2}^x = b_2)}{f_{s_3}^x(b_3|B_{s_1}^x = b_1)} \mathrm{d}b_2 \\
&= \frac{\phi(b_2; b_1, s_2 - s_1)\phi(b_3; b_2, s_3 - s_2)}{\phi(b_3; b_1, s_3 - s_1)} \mathrm{d}b_2 \\
&= \phi\Big(b_2; \frac{(s_3 - s_2)b_1 + (s_2 - s_1)b_3}{s_3 - s_1}, \frac{(s_2 - s_1)(s_3 - s_2)}{s_3 - s_1}\Big) \mathrm{d}b_2
\end{aligned}
$$

where $\phi(u; m, \sigma^2)$ is density of normal distribution $N(m, \sigma^2)$. Thus sampling from Brownian bridge measure is very easy.

Brownian bridge simulation:

INPUT:

$$s_1, s_2, s_3, b_1, b_3$$

1. Sample $U \sim N\big(\frac{(s_3-s_2)b_1+(s_2-s_1)b_3}{s_3-s_1}, \frac{(s_2-s_1)(s_3-s_2)}{s_3-s_1}\big)$
2. Set $b_2 = u$

OUTPUT:

$$b_2$$

3.2 Brownian bridge simulation conditioned that it does not cross fixed level r

Let $0 \leq s_1 < s_2 < s_3 \leq T$. Here we are interested in simulation of B_x at s_2 conditioned on $A := \{B^x_{s_1} = b_1, L^r_{s_1} = l_1 = L^r_{s_2} = L^r_{s_3}, B^x_{s_3} = b_3\}$. We are only interested in the case when the Brownian motion does not hit level $y = l_1$. Here Bessel bridge process can be used to perform simulations.

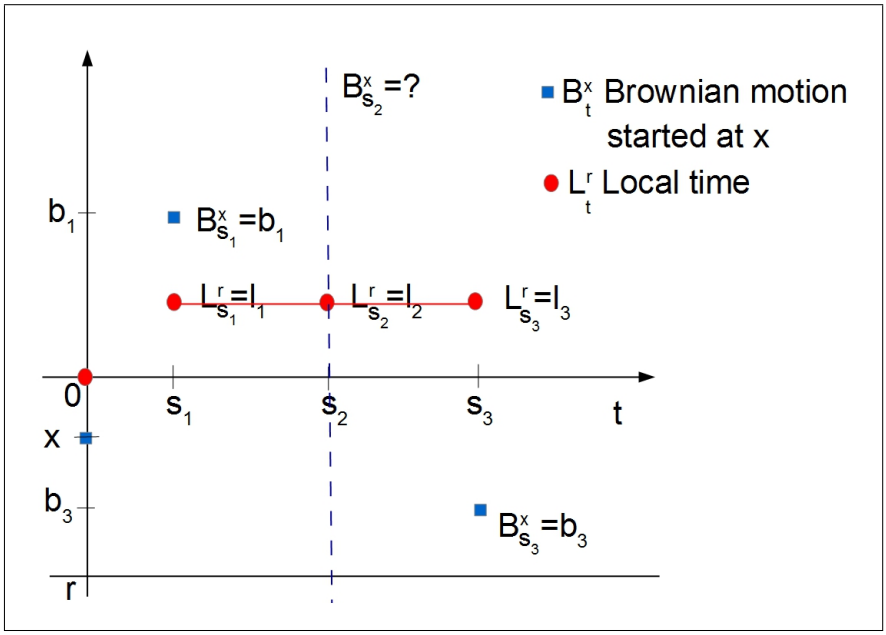


Fig. 2 Unveiled information about Brownian bridge conditioned that it does not hit fixed level r (defining conditions).

3.3 "Completing the right end of joint (B^x, L^r)-bridge" simulation: $L^r_{s_2}$ induced by $\mathbb{BB}^{x,b_1,l_1,b_2}_{s_1,s_2}$

Let $0 \leq s_1 < s_2 \leq T$. Local time process is defined and discussed in Section 2.2. Here we are particularly interested in the regular conditional probability of local time at time s_2 given values of $B_{s_1}, L^r_{s_1}, B_{s_2}$. By "completing the right end of joint (B^x, L^r)-bridge" simulation we mean sampling $L^r_{s_2}$ given $A := \{B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_2} = b_2\}$. The measure induced by B^x conditioned on A on $C([0,T])$ is denoted by $\mathbb{BB}^{x,b_1,l_1,b_2}_{s_1,s_2}$.

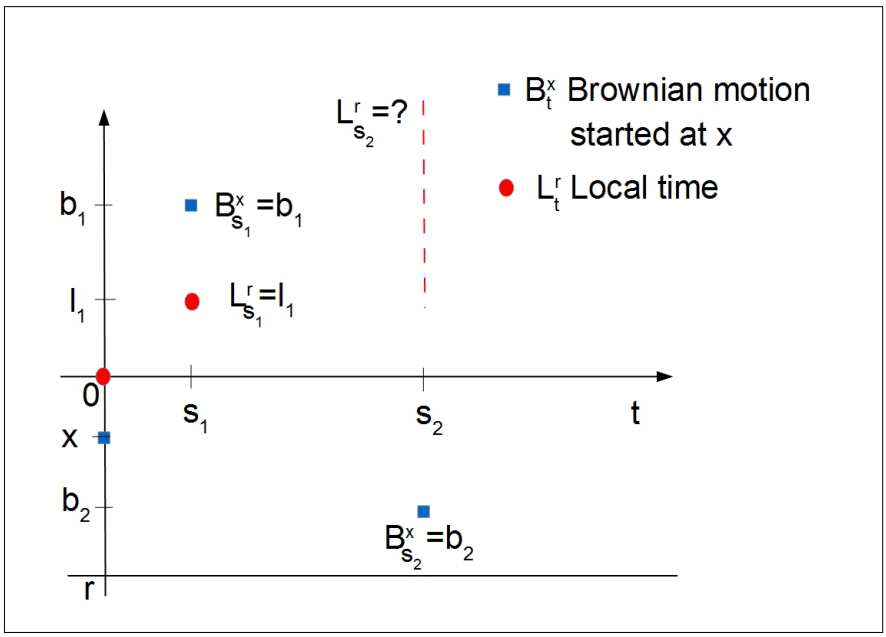


Fig. 3 Unveiled information about incomplete joint (B^x, L^r)-bridge (defining conditions).

This simulation is an integral part of Alg.1 and Alg.2 of Chapter 4. Using formulas (2.6) and (2.7) and Remark 1 we obtain

For $l_2 > l_!$

$$
\begin{aligned}
&P(L^r_{s_2} \in \mathrm{d}l_2 | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_2} = b_2) \\
&= \frac{f^{x,r}_{s_2}(b_2, l_2 | B^x_{s_1} = b_1, L^r_{s_1} = l_1)}{f^{x,r}_{s_2}(b_2 | B^x_{s_1} = b_1, L^r_{s_1} = l_1)} \mathrm{d}l_2 \\
&= \frac{f^{b_1,r}_{s_2-s_1}(b_2, l_2 - l_1)}{\phi(b_2; b_1, s_2 - s_1)} \mathrm{d}l_2 \\
&= \frac{1}{s_2 - s_1}((l_2 - l_1) + |b_2 - r| + |r - b_1|) e^{\frac{(b_2-b_1)^2 - ((l_2-l_1)+|b_2-r|+|r-b_1|)^2}{2(s_2-s_1)}} \mathrm{d}l_2
\end{aligned}
$$

Substituting $l := (l_2 - l_1) + |b_2 - r| + |r - b_1|$ we obtain

$$= \frac{l}{s_2 - s_1} e^{-\frac{l^2}{2(s_2-s_1)}} e^{\frac{(b_2-b_1)^2}{2(s_2-s_1)}} \mathrm{d}l$$

Let $h(u) := \frac{u}{s_2-s_1} e^{-\frac{u^2}{2(s_2-s_1)}}$ for $u \geq 0$ and note that h is the density of Rayleigh distribution with the scale parameter $s_2 - s_1$. Then l has density proportional to the density of Rayleigh distribution with the parameter $(s_2 - s_1)$ truncated to $(|b_2 - r| + |r - b_1|, \infty)$ since $l_2 \in (l_1, \infty)$.

Here we have the formula for the probability that the local time does not increase on $[s_1, s_2]$.

$$\begin{aligned} & P(L^r_{s_2} = l_1 | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_2} = b_2) \\ & = \frac{f^{b_1,r,const}_{s_2-s_1}(b_2)}{\phi(b_2; b_1, s_2 - s_1)} \\ & = 1 - e^{\frac{2((b_2-r)(r-b_1)-|b_2-r||r-b_1|)}{2(s_2-s_1)}} \end{aligned}$$

Finally we arrive at

"Completing the right end of joint (B^x, L^r)-bridge" simulation:

INPUT:

$$s_1, s_2, b_1, l_1, b_2$$

1. If $(b_2 - r)(r - b_1) - |b_2 - r||r - b_1| = 0$ proceed to 2.

   Otherwise

   Sample $U \sim U(0, 1)$

   If $u \leq 1 - e^{\frac{2((b_2-r)(r-b_1)-|b_2-r||r-b_1|)}{2(s_2-s_1)}}$ set $l_2 = l_1$ and finish here

   Otherwise proceed to 2.

2. Sample $U \propto$ [Rayleigh$(s_2 - s_1)$ truncated to $(|b_2 - r| + |r - b_1|, \infty)$]
   Set $l_2 = u + l_1 - |b_2 - r| - |r - b_1|$

OUTPUT:

l_2

3.4 Simulation of "L^r component of joint (B^x, L^r)-bridge": $L^r_{s_2}$ induced by $\mathbb{BB}^{x,b_1,l_1,b_2,b_3,l_3}_{s_1,s_2,s_3}$

Let $0 \leq s_1 < s_2 < s_3 \leq T$. Here we are interested in the local time at the level r at time s_2 given values of $B^x_{s_1}, L^r_{s_1}, B^x_{s_2}, B^x_{s_3}, L^r_{s_3}$. Note that the value of B_x at s_2 has been already "revealed". We show how to simulate $L^r_{s_2}$ of B^x given $A := \{B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_2} = b_2, B^x_{s_3} = b_3, L^r_{s_3} = l_3\}$. Here if $l_1 = l_3$ we only consider A such that Brownian motion does not hit level $y = l_1$. The measure induced by B^x conditioned on A on $C([0,T])$ is denoted by $\mathbb{BB}^{x,b_1,l_1,b_2,b_3,l_3}_{s_1,s_2,s_3}$.

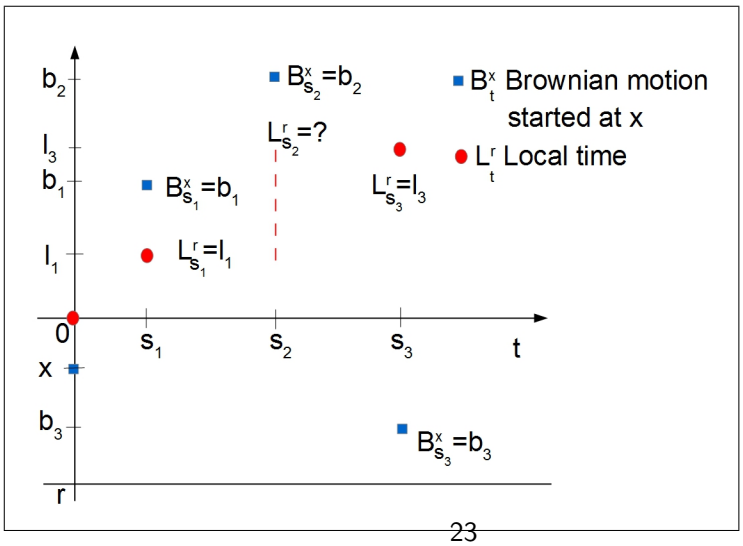


Fig. 4 Unveiled information about "L^r component of joint (B^x, L^r)-bridge" (defining conditions).

"L^r component of joint (B^x, L^r)-bridge" simulation constitutes an inherent part of Alg.2 of Chapter 4.

First note that if $l_1 = l_3$ then $l_1 = l_2 = l_3$ a.s. Thus we now consider $l_1 \neq l_3$ and here we have three distinctive cases: $l_1 = l_2 \neq l_3$, $l_1 \neq l_2 = l_3$ and $l_2 \in (l_1, l_3)$. We consider the first one only if $\mathbb{1}_{b_1 \leq r}\mathbb{1}_{b_2 \leq r} = 1$ or $\mathbb{1}_{b_1 \geq r}\mathbb{1}_{b_2 \geq r} = 1$ and similarly we consider the second one only if $\mathbb{1}_{b_2 \leq r}\mathbb{1}_{b_3 \leq r} = 1$ or $\mathbb{1}_{b_2 \geq r}\mathbb{1}_{b_3 \geq r} = 1$. Let

$$A_1 := f^{x,r,const}_{s_2}(b_2|B^x_{s_1} = b_1, L^r_{s_1} = l_1) f^{x,r}_{s_3}(b_3, l_3|B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_2} = b_2, L^r_{s_2} = l_1)$$
$$= f^{x,r,const}_{s_2}(b_2|B^x_{s_1} = b_1, L^r_{s_1} = l_1) f^{x,r}_{s_3}(b_3, l_3|B^x_{s_2} = b_2, L^r_{s_2} = l_1)$$
$$= f^{b_1,r,const}_{s_2-s_1}(b_2) f^{b_2,r}_{s_3-s_2}(b_3, l_3 - l_1)$$
$$A_2 := f^{x,r}_{s_2}(b_2, l_3|B^x_{s_1} = b_1, L^r_{s_1} = l_1) f^{x,r,const}_{s_3}(b_3|B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_2} = b_2, L^r_{s_2} = l_3)$$
$$= f^{b_1,r}_{s_2-s_1}(b_2, l_3 - l_1) f^{b_2,r,const}_{s_3-s_2}(b_3)$$
$$A_3(l_2) := f^{x,r}_{s_2}(b_2, l_2|B^x_{s_1} = b_1, L^r_{s_1} = l_1) f^{x,r}_{s_3}(b_3, l_3|B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_2} = b_2, L^r_{s_2} = l_2)$$
$$= f^{b_1,r}_{s_2-s_1}(b_2, l_2 - l_1) f^{b_2,r}_{s_3-s_2}(b_3, l_3 - l_2).$$

Observe that

$$A_4 := \int_{l_2 \in (l_1, l_3)} A_3(l_2) dl_2$$
$$= \int_{l_2 \in (l_1, l_3)} f^{b_1,r}_{s_2-s_1}(b_2, l_2 - l_1) f^{b_2,r}_{s_3-s_2}(b_3, l_3 - l_2) dl_2$$
$$= \int_{l_2 \in (l_1, l_3)} \frac{1}{(s_2 - s_1)\sqrt{2\pi(s_2 - s_1)}}(l_2 - l_1 + |b_2 - r| + |r - b_1|) e^{-\frac{(l_2 - l_1 + |b_2 - r| + |r - b_1|)^2}{2(s_2 - s_1)}}$$
$$\times \frac{1}{(s_3 - s_2)\sqrt{2\pi(s_3 - s_2)}}(l_3 - l_2 + |b_3 - r| + |r - b_2|) e^{-\frac{(l_3 - l_2 + |b_3 - r| + |r - b_2|)^2}{2(s_3 - s_2)}} dl_2$$
$$= k_1[\sigma(t_1 e^{-\frac{t_1^2}{2}} - t_2 e^{-\frac{t_2^2}{2}}) + k_2(\Phi(t_2) - \Phi(t_1))]$$

where

$$\alpha := -l_1 + |b_2 - r| + |r - b_1|$$

$$\beta := l_3 + |b_3 - r| + |r - b_2|$$

$$k_1 := (2\pi)^{-1}(s_2 - s_1)^{-\frac{3}{2}}(s_3 - s_2)^{-\frac{3}{2}} e^{\frac{(\alpha+\beta)^2}{2(s_3-s_1)}}$$

$$\mu := \frac{\beta(s_2 - s_1) - \alpha(s_3 - s_2)}{(s_3 - s_1)}$$

$$\sigma^2 := \frac{(s_2 - s_1)(s_3 - s_2)}{(s_3 - s_1)}$$

$$k_2 := -\sigma + (\beta - \alpha - 2\mu) + \frac{1}{\sigma}(\alpha\beta - \mu^2 + \mu(\beta - \alpha))$$

$$t_1 := \frac{l_1 - \mu}{\sigma}$$

$$t_2 := \frac{l_2 - \mu}{\sigma}$$

$\Phi -$ cumulative distribution function of normal distribution

Then

$$P(L^r_{s_2} = l_1 | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_2} = b_2, B^x_{s_3} = b_3, L^r_{s_3} = l_3)$$
$$= \frac{A_1}{A_1 + A_2 + A_4}$$
$$P(L^r_{s_2} = l_3 | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_2} = b_2, B^x_{s_3} = b_3, L^r_{s_3} = l_3)$$
$$= \frac{A_2}{A_1 + A_2 + A_4}$$
$$P(L^r_{s_2} \in \mathrm{d}l_2 | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_2} = b_2, B^x_{s_3} = b_3, L^r_{s_3} = l_3)$$
$$= \frac{A_3}{A_1 + A_2 + A_4} dl_2.$$

The above computations allow us to formulate the following procedure

Simulation of "L^r component of joint (B^x, L^r)-bridge"

INPUT:

$$s_1, s_2, s_3, b_1, l_1, b_2, b_3, l_3$$

If $l_1 = l_3$ then

If $\mathbb{1}_{b_1<r}\mathbb{1}_{b_2<r}\mathbb{1}_{b_3<r} = 1$ or $\mathbb{1}_{b_1>r}\mathbb{1}_{b_2>r}\mathbb{1}_{b_3>r} = 1$ then set $l_2 = l_1$ and finish here

Otherwise print "error" and finish here

Otherwise (when $l_1 \neq l_3$)

1. If $\mathbb{1}_{b_1\geq r}\mathbb{1}_{b_2\leq r} = 1$ or $\mathbb{1}_{b_1\leq r}\mathbb{1}_{b_2\geq r} = 1$ then set $A_1 = 0$

   Otherwise compute A_1 and A_4

   If $\mathbb{1}_{b_2\geq r}\mathbb{1}_{b_3\leq r} = 1$ or $\mathbb{1}_{b_2\leq r}\mathbb{1}_{b_3\geq r} = 1$ then set $A_2 = 0$

   Otherwise compute A_2

2. If $A_1 = 0$ and $A_2 = 0$ proceed to (3.)

   Otherwise

   Sample $U_1 \sim U(0, 1)$

   If $u_1 \leq \frac{A_1}{A_1+A_2+A_4}$ set $l_2 = l_1$ and finish here

   If $\frac{A_1}{A_1+A_2+A_4} < u_1 \leq \frac{A_1+A_2}{A_1+A_2+A_4}$ set $l_2 = l_3$ and finish here

   Otherwise proceed to (3.)

3. Sample (e.g. by using rejection sampling) U_2 with density

$$h(u_2) \propto \mathbb{1}_{(l_1,l_3)}(u_2) f^{b_1,r}_{s_2-s_1}(b_2, u_2 - l_1) f^{b_2,r}_{s_3-s_2}(b_3, l_3 - u_2)$$

Set $l_2 = u_2$

OUTPUT:

l_2

3.5 (B^x, L^r)-bridge simulation: $(B^x_{s_2}, L^r_{s_2}) \sim \mathbb{BB}^{x,b_1,l_1,b_3,l_3}_{s_1,s_3}$

Consider times $0 \leq s_1 < s_2 < s_3 \leq T$. What we mean here by (B^x_s, L^r_s)-bridge simulation is sampling from joint distribution of $B^x_{s_2}$ - Brownian motion started at x and its local time at level r, $L^r_{s_2}$, conditioned on $A := \{B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_3} = b_3, L^r_{s_3} = l_3\}$. The measure induced by B^x conditioned on A on $C([0,T])$ is denoted by $\mathbb{BB}^{x,b_1,l_1,b_3,l_3}_{s_1,s_3}$.

The procedure for sampling (B^x, L^r)-bridges is crucial for the algorithms Alg.2 and Alg.3 and also allows to refine skeletons $(B^x_i, L^r_i)_{i\in I}$, obtained from Alg.1, Alg.2 and Alg.3 **after** the path has been accepted as we see in Chapter 4.

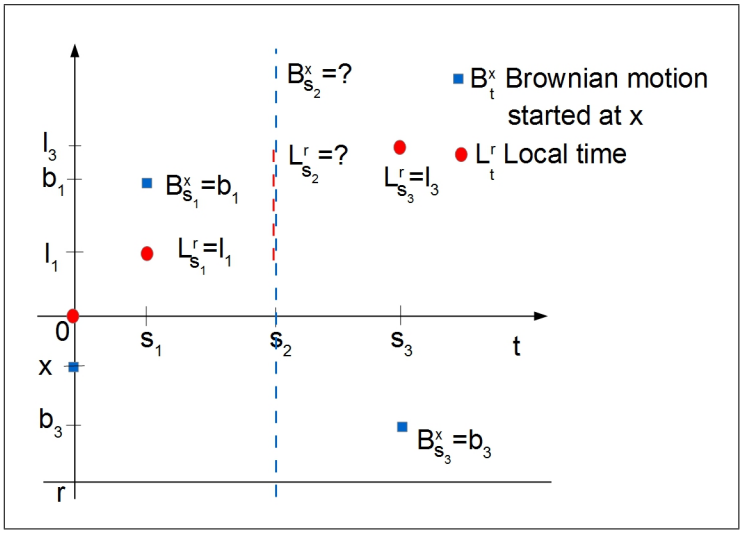


Fig. 5 Unveiled information about "$(B^x_{s_2}, L^r_{s_2}) \sim \mathbb{BB}^{x,b_1,l_1,b_3,l_3}_{s_1,s_3}$" (defining conditions).

First we consider the situation where $l_1 \neq l_3$. Similarly to Section 3.4 we need to consider here three cases and we introduce some notation.

1. $l_1 = l_2$ but $l_1 \neq l_3$. Let $A_1(b_2)$ such that

$$A_1(b_2)db_2 := P(B^x_{s_2} \in \mathrm{d}b_2, L^r_{s_2} = l_1 | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_3} = b_3, L^r_{s_3} = l_3)$$
$$p_1 := \int_{-\infty}^{\infty} A_1(b_2)db_2 = P(L^r_{s_2} = l_1 | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_3} = b_3, L^r_{s_3} = l_3)$$

2. $l_2 = l_3$ but $l_1 \neq l_3$. Let $A_1(b_2)$ such that

$$A_2(b_2)db_2 := P(B^x_{s_2} \in \mathrm{d}b_2, L^r_{s_2} = l_3 | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_3} = b_3, L^r_{s_3} = l_3)$$
$$p_2 := \int_{-\infty}^{\infty} A_2(b_2)db_2 = P(L^r_{s_2} = l_3 | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_3} = b_3, L^r_{s_3} = l_3)$$

3. $l_2 \in (l_1, l_3)$ and $l_1 \neq l_3$. Note that local time is an increasing process.

$$A_3(b_2, l_2)db_2dl_2 := P(B^x_{s_2} \in \mathrm{d}b_2, L^r_{s_2} \in \mathrm{d}l_2 | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_3} = b_3, L^r_{s_3} = l_3)$$
$$p_3 := \int_{\mathbb{R}} \int_{(l_1,l_3)} A_3(b_2, l_2)dl_2db_2$$
$$= P(L^r_{s_2} \in (l_1, l_3) | B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_3} = b_3, L^r_{s_3} = l_3).$$

Observe that $p_1+p_2+p_3=1$. Three cases and mixed probability distribution arise as a result of the fact that local time distribution has a point mass. We introduce p_1, p_2 and p_3 so that we can use it to split simulation into two steps. First we determine the case and then conditioned on the case we sample the value of $B^x_{s_2}$ (or in case (3.) of both: $B^x_{s_2}$ and $L^r_{s_2}$). Observe that

$$\begin{aligned}
A_1 &= \frac{f^{x,r,const}_{s_2}(b_2|B^x_{s_1}=b_1, L^r_{s_1}=l_1) f^{x,r}_{s_3}(b_3,l_3|B^x_{s_1}=b_1, L^r_{s_1}=l_1, B^x_{s_2}=b_2, L^r_{s_2}=l_1)}{f^{x,r}_{s_3}(b_3,l_3|B^x_{s_1}=b_1, L^r_{s_1}=l_1)} \\
&= \frac{f^{b_1,r,const}_{s_2-s_1}(b_2) f^{b_2,r}_{s_3-s_2}(b_3, l_3-l_1)}{f^{b_1,r}_{s_3-s_1}(b_3,l_3-l_1)} \\
p_1 &= \frac{\int_{-\infty}^{\infty} f^{b_1,r,const}_{s_2-s_1}(b_2) f^{b_2,r}_{s_3-s_2}(b_3, l_3-l_1)}{f^{b_1,r}_{s_3-s_1}(b_3,l_3-l_1)}
\end{aligned}$$

$$\begin{aligned}
A_2 &= \frac{f^{x,r}_{s_2}(b_2,l_3|B^x_{s_1}=b_1, L^r_{s_1}=l_1) f^{x,r,const}_{s_3}(b_3|B^x_{s_1}=b_1, L^r_{s_1}=l_1, B^x_{s_2}=b_2, L^r_{s_2}=l_3)}{f^{x,r}_{s_3}(b_3,l_3|B^x_{s_1}=b_1, L^r_{s_1}=l_1)} \\
&= \frac{f^{b_1,r}_{s_2-s_1}(b_2, l_3-l_1) f^{b_2,r,const}_{s_3-s_2}(b_3)}{f^{b_1,r}_{s_3-s_1}(b_3,l_3-l_1)} \\
p_2 &= \frac{\int_{-\infty}^{\infty} f^{b_1,r}_{s_2-s_1}(b_2, l_3-l_1) f^{b_2,r,const}_{s_3-s_2}(b_3)}{f^{b_1,r}_{s_3-s_1}(b_3,l_3-l_1)}
\end{aligned}$$

$$\begin{aligned}
A_3 =& P(B^x_{s_2}\in \mathrm{d}b_2, L^r_{s_2}\in \mathrm{d}l_2 | B^x_{s_1}=b_1, L^r_{s_1}=l_1, B^x_{s_3}=b_3, L^r_{s_3}=l_3) \\
&= \frac{f^{x,r}_{s_2}(b_2,l_2|B^x_{s_1}=b_1, L^r_{s_1}=l_1) f^{x,r}_{s_3}(b_3,l_3|B^x_{s_1}=b_1, L^r_{s_1}=l_1, B^x_{s_2}=b_2, L^r_{s_2}=l_2)}{f^{x,r}_{s_3}(b_3,l_3|B^x_{s_1}=b_1, L^r_{s_1}=l_1)} \\
&= \frac{f^{b_1,r}_{s_2-s_1}(b_2, l_2-l_1) f^{b_2,r}_{s_3-s_2}(b_3, l_3-l_2)}{f^{b_1,r}_{s_3-s_1}(b_3,l_3-l_1)} \mathrm{d}b_2 \mathrm{d}l_2 \\
&= \frac{\frac{1}{(s_2-s_1)\sqrt{2\pi(s_2-s_1)}}(l_2-l_1+|b_2-r|+|r-b_1|) e^{-\frac{(l_2-l_1+|b_2-r|+|r-b_1|)^2}{2(s_2-s_1)}}}{f^{b_1,r}_{s_3-s_1}(b_3,l_3-l_1)} \\
&\times \frac{1}{(s_3-s_2)\sqrt{2\pi(s_3-s_2)}}(l_3-l_2+|b_3-r|+|r-b_2|) e^{-\frac{(l_3-l_2+|b_3-r|+|r-b_2|)^2}{2(s_3-s_2)}}
\end{aligned} \tag{3.1}$$

Note in (3.1) the symmetry in b_2 about $b_2=r$. Assume that $b_2>r$ then by

substituting

$$u := l_2 - l_1 + |b_2 - r| + |r - b_1|$$

$$v := l_3 - l_2 + |b_3 - r| + |r - b_2|$$

we further have

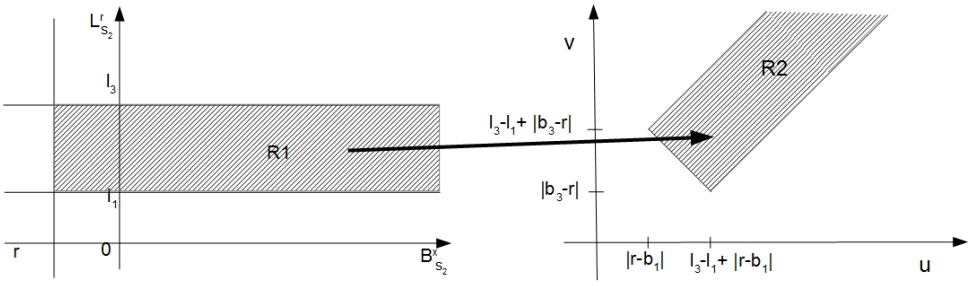
$$A_3 = c_{uv} u e^{-\frac{u^2}{2(s_2 - s_1)}} v e^{-\frac{v^2}{2(s_3 - s_2)}} \mathrm{d}u\mathrm{d}v$$

where $c_{uv}^{-1} = 4\pi(s_2 - s_1)(s_3 - s_2)\sqrt{(s_2 - s_1)(s_3 - s_2)} f_{s_3 - s_1}^{b_1, r}(b_3, l_3 - l_1)$. It is important to note that the distribution of $(B_{s_2}^x, L_{s_2}^r | B_{s_1}^x = b_1, L_{s_1}^r = l_1, B_{s_3}^x = b_3, L_{s_3}^r = l_3)$ is a.s. equal to 0 outside $\mathbb{R} \times [l_1, l_3]$. Under the linear transformation $(b_2, l_2) \to (u, v)$ the region $R_1 := [r, \infty) \times [l_1, l_3]$ is mapped to the region R_2 bounded by the following lines:

$$v = u + (l_1 - l_3 - |r - b_1| + |b_3 - r|)$$

$$v = u - (l_1 - l_3 - |r - b_1| + |b_3 - r|)$$

$$v = -u + (l_3 - l_1 + |r - b_1| + |b_3 - r|).$$



If $l_1 = l_3$

Use procedure from Section 3.2

Hence the procedure to sample $B^x_{s_2}$ and $L^r_{s_2}$ conditioned on $\{B^x_{s_1} = b_1, L^r_{s_1} = l_1, B^x_{s_3} = b_3, L^r_{s_3} = l_3\}$ is as follows:

$(\boldsymbol{B^x}, \boldsymbol{L^r})$-bridge simulation

INPUT:

$$s_1, s_2, s_3, b_1, l_1, b_3, l_3$$

If $l_1 \neq l_3$

1. Sample $Z_1 \sim U(0,1)$
2. Compute p_1. If $z_1 > p_1$ proceed to (3.) otherwise

   Set $l_2 = l_1$

   Sample $Z_2 \sim h(u) \propto A_1(u)$

   Set $b_2 = z_2$ and finish here
3. Compute p_2. If $z_1 > p_1 + p_2$ proceed to (4.) otherwise

   Set $l_2 = l_3$

   Sample $Z_2 \sim h(u) \propto A_2(u)$

   Set $b_2 = z_2$ and finish here
4. Sample $(U, V) \sim h(u,v) \propto u e^{-\frac{u^2}{2(s_2 - s_1)}} v e^{-\frac{v^2}{2(s_3 - s_2)}}$ on R_2

Set $l_2 = \frac{u-v+l_3+l_1-|r-b_1|+|b_3-r|}{2}$

Sample $Z \sim \text{Ber}(0.5)$

If $Z = 1$ then set $b_2 = r + \frac{u+v-l_3+l_1-|r-b_1|-|b_3-r|}{2}$

otherwise set $b_2 = r - \frac{u+v-l_3+l_1-|r-b_1|-|b_3-r|}{2}$.

If $l_1 = l_3$ then

If $\mathbb{1}_{b_1<r}\mathbb{1}_{b_2<r}\mathbb{1}_{b_3<r} = 1$ or $\mathbb{1}_{b_1>r}\mathbb{1}_{b_2>r}\mathbb{1}_{b_3>r} = 1$ then

Set $l_2 = l_1$

Sample b_2 as in Section 3.2 and finish here

Otherwise print "error" and finish here

OUTPUT:

b_2, l_2

Chapter 4

Algorithms

In Section 2.1 we described the derivation of EA1 and here we discuss how we can modify this algorithm to become applicable for diffusions with discontinuous drift. Again we are looking at the solutions of one-dimensional SDEs of the form

$$dX_t = \alpha(X_t)dt + dB_t, \quad t \in [0, T], X_t = x. \tag{4.1}$$

where B denotes one-dimensional Brownian motion. Let us start working with the assumptions of EA1 with temporarily omitting the assumption on differentiability of the drift function α. We use the same notation as in Section 2.1.

Set of conditions for Alg.1, Alg.2 and Alg.3 - To be continued

Assumption 1. *There exists non-explosive weak solution to (4.1)*

Assumption 2. *The Girsanov theorem may be applied and the local martingale is a true martingale.*

Assumption 3. *To be modified.*

Assumption 4. *Function* $\exp\{A(u) - \frac{(u-x)^2}{2T}\}$ *is integrable.*

Assumption 5. *Function* $(\alpha^2(u) + \alpha'(u))\mathbb{1}_{u \notin D}$ *is bounded below*

Set D is defined in Assumption 3.1. Recall that we denote by $\mathbb{W}$ the measure induced by Brownian motion started at x on $(C, \mathcal{C})$ and by $\mathbb{Q}$ the measure induced by the diffusion X. Under Assumption 2 we may apply Girsanov transformation to obtain the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{W}$.

$$\frac{d\mathbb{Q}}{d\mathbb{W}} = \exp\left\{ \int_0^T \alpha(X_t) dX_t - \frac{1}{2} \int_0^T \alpha^2(X_t) dt \right\}$$

Define function $A : \mathbb{R} \to \mathbb{R}$ in the same way as for EA1, namely $A(u) := \int_0^u \alpha(y) dy$. Here is the point where we would like to proceed by applying Itô's lemma but it is not possible as α is not differentiable. Let us introduce the following assumption instead.

Assumption 3.1. *Let $r_1 < r_2 < \cdots < r_n$ be real numbers, and define $D = \{r_1, r_2, \cdots, r_n\}$. Assume that the drift function $\alpha : \mathbb{R} \to \mathbb{R}$ is continuous on $\mathbb{R} \setminus D$ and α' exists and is continuous on $\mathbb{R} \setminus D$, and the limits*

$$\alpha'(r_k \pm) := \lim_{x \to r_k \pm} \alpha'(x)$$

exist and are finite.

We begin referring to local time process here and more details can be found in Section 2.2. Referring to Problem 6.24 in [Kar91] (pp. 215) we note that the following Remark 2 results from the Assumption 3.1 and Theorem 2.6.

Remark 2. Under the Assumption 3.1 A is the difference of two convex functions and

$$A(B_t) = A(x) + \int_0^t \alpha(B_s) dB_s + \frac{1}{2} \int_0^t \alpha'(B_s) ds + \\ + \frac{1}{2} \sum_{k=1}^n L_t^{r_k} [\alpha(r_k+) - \alpha(r_k-)], \text{ a.s.} \quad t \in [0, T]$$

By Remark 2 we obtain that

$$\frac{d\mathbb{Q}}{d\mathbb{W}} = \exp\{A(X_T) - A(x) - \frac{1}{2}\int_0^T (\alpha'(X_t) + \alpha^2(X_t))dt$$
$$- \frac{1}{2}\sum_{k=1}^{n} L_T^{r_k}(X)[\alpha(r_k+) - \alpha(r_k-)]\} \text{ a.s.} \tag{4.2}$$

Occurrence of local time at multiple levels in formula (4.2) constitutes a great challenge and it is not known how to deal with it. The algorithms that we present address the problem in case the drift function α has one point of discontinuity. We denote this point by r. Then the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{W}$ is equal to

$$\frac{d\mathbb{Q}}{d\mathbb{W}} = \exp\{A(X_T) - A(x) - \frac{1}{2}\int_0^T (\alpha'(X_t) + \alpha^2(X_t))dt$$
$$- \frac{1}{2}L_T^{r}(X)[\alpha(r+) - \alpha(r-)]\} \text{ a.s.} \tag{4.3}$$

Now the big question is how to use the knowledge of (4.3) for simulation purpose. We suggest three different answers in the following Sections and we call the new algorithms Alg.1, Alg.2 and Alg.3.

4.1 Alg.1

As we discussed in previous sections EA1 is based on rejection sampling performed in special way. Let us discuss now what are the issues about adapting the ideas of EA1 directly. There are three main questions:

- Shall we sample candidate paths from biased Brownian motion or there exists better candidate process?
- Is the Radon-Nikodym derivative of diffusion measure with respect to chosen measure bounded? As it is required to perform rejection sampling.

- Can we construct an auxiliary event whose probability is equal to the acceptance probability of a path and it requires only finite information about a path?

Let's see what happens when we choose biased Brownian motion to be our candidate process. Applying (2.4) and (4.3) we arrive at

$$\frac{d\mathbb{Q}}{d\mathbb{Z}} = \frac{d\mathbb{Q}}{d\mathbb{W}}\frac{d\mathbb{W}}{d\mathbb{Z}} \propto \exp\Big\{ -\frac{1}{2}\int_0^T (\alpha^2(X_t) + \alpha'(X_t))dt - \frac{1}{2}L_T^r(X)[\alpha(r+) - \alpha(r-)]\Big\}.$$

Observe that Assumption 5 assures that $\exp\Big\{ -\frac{1}{2}\int_0^T (\alpha^2(X_t) + \alpha'(X_t))dt\Big\}$ is bounded. What about

$$\exp\Big\{ -\frac{1}{2}L_T^r(X)[\alpha(r+) - \alpha(r-)]\Big\}$$

then?

Recall that local time is a non-negative process. Thus under the following assumption

Assumption 3.2. $\alpha(r+) - \alpha(r-) \geq 0.$

we obtain

$$\frac{d\mathbb{Q}}{d\mathbb{Z}} \propto \exp\Big\{ -\int_0^T \phi(B_t)dt - \frac{1}{2}L_T^r(X)[\alpha(r+) - \alpha(r-)]\Big\} \leq 1; \quad \mathbb{Z} - a.s.$$

and it is great news as we have established boundedness of Radon-Nikodym derivative now which is essential for the rejection sampling. In EA1 rejection sampling is performed retrospectively, i.e. first auxiliary decision (acceptance or rejection) variable is sampled and then values of the candidate process at necessary time points. Here we also need values of B at some random times but **jointly** with the value of local time at level r at time T. Two different methods can be introduced to achieve it which lead to construction of Alg.1 and Alg.2. Let us concentrate now

on Alg.1. We propose to split the rejection sampling into two steps. In the first one, which we refer to as Step 1 we perform rejection sampling with acceptance probability p_1 as follows

$$p_1(X) \propto \exp\{-\int_0^T \phi(X_t)dt\}.$$

If the path was not rejected then in Step 2 we again use rejection sampling with the acceptance probability p_2

$$p_2(X) \propto \exp\{-L_T^r(X)\frac{\alpha(r^+)-\alpha(r^-)}{2}\}.$$

where $L_T^r(X)$ **is sampled conditioned on the revealed values of** X **in Step 1**.

Let us now put together what we discussed above and present the algorithm Alg.1. It requires Assumptions 1, 2, 3.1, 3.2, 4 and 5. The following algorithm returns samples from finite-dimensional distribution of the diffusion X together with its local times.

Alg.1

INPUT:

$\alpha(u) : \mathbb{R} \to \mathbb{R}$ drift function

$\{0 = t^0, t^1, \cdots, t^n, t^{n+1} = T\}$ time instances of interest

$X_0 = x$ initial point

I 1. Define

(a) $A : \mathbb{R} \to \mathbb{R}$, $A(u) := \int_0^u \alpha(y)dy$

(b) $k = \inf_{u\in\mathbb{R}} \frac{\alpha^2(u)+\alpha'(u)}{2}$

(c) $\phi : \mathbb{R} \to \mathbb{R}$, $\phi(u) := \frac{\alpha^2(u)+\alpha'(u)}{2} - k$

2. Find the upper bound M of ϕ.

II Generate realisation of X at T and at random times in $(0, T)$:

1. Sample $X_T \sim h(u) \propto \exp\{A(u) - \frac{(u-x)^2}{2T}\}$

2. Sample an auxiliary planar Poisson process Ψ with intensity 1

(i) Sample $k \sim$ Poiss(TM)

(ii) Sample $U_1, \cdots, U_k$ i.i.d., $U_1 \sim$ Uni$((0, T) \times (0, M))$

(iii) Sort $\{u_1, \cdots, u_k\}$ in increasing order of their first coordinate τ to get $((\tau_1, \psi_1), \cdots, (\tau_k, \psi_k))$

3. Use Ψ to perform first rejection sampling:

(a) Sample $(X_{\tau_1}, \cdots, X_{\tau_k})$ from the Brownian bridge measure $\mathbb{W}_{0,T}^{x,x_T}$

(b) Compute $\phi(X_{\tau_i})$ for $i \in \{1, \cdots, k\}$

(c) If $\phi(X_{\tau_i}) < \psi_{\tau_i}$ for each $i \in \{1, \cdots, k\}$ then proceed to **III**. Otherwise return to **II**.**1**

III Generate realisation of X at $(t^1, \cdots, t^n)$ conditioned on values of X at $(0, \tau_1, \cdots, \tau_k, T)$:

(a) Sort in increasing order the set $\{0, \tau_1, \cdots, \tau_k, t^1, \cdots, t^n, T\}$ to get $(0, t_1, \cdots, t_{k+n}, T)$

(b) Repeat n times

(i) Pick $i \in \{1, \cdots, k+n\}$ such that X_{t_i} has not been sampled yet

(ii) Find $t_L := \max_{\{0,t_1,\cdots,t_{k+n},T\}}\{t_j$ such that X_{t_j} has been already revealed and $t_j < t_i\}$

(iii) Find $t_U := \min_{\{0,t_1,\cdots,t_{k+n},T\}}\{t_j$ such that X_{t_j} has been already revealed and $t_j > t_i\}$

(iv) Sample X_{t_i} from the Brownian bridge measure $\mathbb{BB}^{x_L,x_U}_{t_L,t_U}$ [See Section 3.1]

IV Generate realisation of $L^r(X)$ at $(0, t_1, \cdots, t_{k+n}, T)$ conditioned on $(x, X_{t_1}, \cdots, X_{t_{k+n}}, X_T)$

Set $L^r(X)_0 = 0$

For $i \in (t_1, \cdots, t_{k+n}, T)$ perform recursively

Sample $l \sim \mathbb{BB}^{x,b_{t_{1-1}},l_{t_{1-1}},b_{t_i}}_{t_{i-1},t_i}$ [See Section 3.3]

Set $L^r(X)_{t_i} = L^r(X)_{t_{i-1}} + l$

V Perform rejection sampling using L^r_T with acceptance probability p_2

If the path is accepted return

$$((0,x,0),(t_1,X_{t_1},L^r(X)_{t_1}),\cdots,(t_{k+n},X_{t_{k+n}},L^r(X)_{t_{k+n}}),(T,X_T,L^r(X)_T)$$

Otherwise start again at II.1

OUTPUT:

$$((0,x,0),(t_1,X_{t_1},L^r(X)_{t_1}),\cdots,(t_{k+n},X_{t_{k+n}},L^r(X)_{t_{k+n}}),(T,X_T,L^r(X)_T)$$

4.2 Alg.2

Alg.2 requires the same assumptions as Alg.1. However it involves different order of simulations and possibility to sample from joint local time and Brownian motion bridges.

Alg.2

INPUT:

$\alpha(u) : \mathbb{R} \to \mathbb{R}$ drift function

$\{0 = t^0, t^1, \cdots, t^n, t^{n+1} = T\}$ time instances of interest

$X_0 = x$ initial point

I 1. Define

(a) $A : \mathbb{R} \to \mathbb{R}$, $A(u) := \int_0^u \alpha(y)dy$

(b) $k = \inf_{u \in \mathbb{R}} \frac{\alpha^2(u)+\alpha'(u)}{2}$

(c) $\phi : \mathbb{R} \to \mathbb{R}$, $\phi(u) := \frac{\alpha^2(u)+\alpha'(u)}{2} - k$

2. Find the upper bound M of ϕ.

II Generate realisation of X at T:

1. Sample $X_T \sim h(u) \propto \exp\{A(u) - \frac{(u-x)^2}{2T}\}$

III Generate realisation of $L^r(\omega)$ at T conditioned on $B_0 = 0, L_0^r = 0$ and $X_T = x_T$

IV Sample an auxiliary planar Poisson process Ψ with intensity 1

(i) Sample $k \sim \text{Poiss}(TM)$

(ii) Sample $U_1, \cdots, U_k$ i.i.d., $U_1 \sim \text{Uni}((0,T) \times (0,M))$

(iii) Sort $\{u_1, \cdots, u_k\}$ in increasing order of their first coordinate τ to get $((\tau_1, \psi_1), \cdots, (\tau_k, \psi_k))$

V Generate B^x jointly with L^r conditioned on values at 0 and T:

(a) Sort in increasing order the set $\{0, \tau_1, \cdots, \tau_k, t^1, \cdots, t^n, T\}$ to get $(0, t_1, \cdots, t_{k+n}, T)$

(b) Repeat $k+n$ times

(i) Pick $i \in \{1, \cdots, k+n\}$ such that X_{t_i} has not been sampled yet

(ii) Find $t_L := \max_{\{0,t_1,\cdots,t_{k+n},T\}}\{t_j$ such that X_{t_j} has been already revealed and $t_j < t_i\}$

(iii) Find $t_U := \min_{\{0,t_1,\cdots,t_{k+n},T\}}\{t_j$ such that X_{t_j} has been already revealed and $t_j > t_i\}$

(iv) Sample $(X_{t_i}, L^r_{t_i})$ using the joint Brownian bridge and local time $\mathbb{BB}^{x,b_L,l_L,b_U,l_U}_{t_L,t_U}$ [See Section 3.5]

VI Use Ψ to perform first rejection sampling:

(a) Compute $\phi(X_{\tau_i})$ for $i \in \{1, \cdots, k\}$

(b) If $\phi(X_{\tau_i}) < \psi_{\tau_i}$ for each $i \in \{1, \cdots, k\}$ then proceed to **VII**. Otherwise return to **II.1**

VII Perform rejection sampling using L^r_T with acceptance probability p_2

If the path is accepted return

$$((0,x,0),(t_1,X_{t_1},L^r(X)_{t_1}),\cdots,(t_{k+n},X_{t_{k+n}},L^r(X)_{t_{k+n}}),(T,X_T,L^r(X)_T)$$

Otherwise start again at II.1

OUTPUT:

$$((0,x,0),(t_1,X_{t_1},L^r(X)_{t_1}),\cdots,(t_{k+n},X_{t_{k+n}},L^r(X)_{t_{k+n}}),(T,X_T,L^r(X)_T)$$

4.3 Alg.3

In Alg.3 we suggest to use another candidate process. Motivation behind this choice is to find a measure such that the Radon-Nikodym derivative of diffusion measure with respect to new measure will not include explicitly any terms containing L^r_X. Otherwise, as we saw in Alg.1 and Alg.2, the usage of the algorithms can become restricted to the case when $\alpha(r+) > \alpha(r-)$ (Assumption 3.2). We define measure $\mathbb{S}$ on $C([0,T])$ with paths starting a.s. at x satisfying

$$\frac{d\mathbb{S}}{d\mathbb{W}} \propto \mathbb{1}_{\{L^r_T(X)>0\}}\frac{g(X_T,L^r_T(X))}{f(X_T,L^r_T(X)} + \mathbb{1}_{\{L^r_T(X)=0\}}\frac{g_0(X_T)}{f_0(X_T)}$$

where functions $f : \mathbb{R} \times (0,\infty) \to \mathbb{R}_+,\ f_0 : \mathbb{R} \to \mathbb{R}_+,\ g : \mathbb{R} \times (0,\infty) \to \mathbb{R}_+$ and $g_0 : \mathbb{R} \to \mathbb{R}_+$ such that

$$
\begin{aligned}
&P(B_T^x \in \mathsf{d}z, L_T^r(B^x) \in \mathsf{d}y) = f(z,y)\ \mathsf{d}z\mathsf{d}y \quad \text{for } y > 0\\
&P(B_T^x \in \mathsf{d}z, L_T^r(B^x = 0)) = f_0(z)\ \mathsf{d}z\\
&g(z,y) := c_g f(z,y) e^{A(z) - y\frac{\alpha(r+)-\alpha(r-)}{2}} \quad \text{for } y > 0\\
&= c_g (s\sqrt{2\pi s})^{-1}(y + |z-r| + |r-x|) e^{-\frac{(y+|z-r|+|r-x|)^2}{2s}} e^{A(z) - y\frac{\alpha(r+)-\alpha(r-)}{2}}\\
&g_0(z) := c_g f_0(z) e^{A(z)}\\
&= c_g (\sqrt{2\pi s}^{-1}) e^{A(z)} (e^{-\frac{(z-x)^2}{2s}} - e^{-\frac{(|z-r|+|r-x|)^2}{2s}})\\
&\text{where } c_g \text{ such that} \quad \int_{\mathbb{R}} \int_{(0,\infty)} g(z,y)\mathsf{d}y\mathsf{d}z + \int_{\mathbb{R}} g_0(z)\mathsf{d}z = 1
\end{aligned}
$$

Let us now show that measure $\mathbb{S}$ is very useful as a candidate measure for rejection sampling. We compute appropriate Radon-Nikodym derivative

$$
\begin{aligned}
\frac{d\mathbb{Q}}{d\mathbb{S}} = \frac{d\mathbb{Q}}{d\mathbb{W}} \frac{d\mathbb{W}}{d\mathbb{S}} &\propto e^{A(X_T) - A(x) - \frac{1}{2}\int_0^T (\alpha'(X_t) + \alpha^2(X_t))dt - \frac{1}{2} L_T^r(X)[\alpha(r+) - \alpha(r-)]}\\
&\times \Big(\mathbb{1}_{\{L_T^r(X) > 0\}} \frac{c_g f(X_T, L_T^r(X)) e^{A(X_T) - L_T^r(X)\frac{\alpha(r+)-\alpha(r-)}{2}}}{f(X_T, L_T^r(X)}\\
&+ \mathbb{1}_{\{L_T^r(X)=0\}} c_g \frac{f_0(X_T) e^{A(X_T)}}{f_0(X_T)} \Big)^{-1}\\
&\propto c_g^{-1} e^{-A(x) - \frac{1}{2}\int_0^T (\alpha'(X_t) + \alpha^2(X_t))dt} \text{ a.s.} \qquad (4.4)
\end{aligned}
$$

From the construction of measure $\mathbb{S}$ we see that if processes generated by measures $\mathbb{S}$ and $\mathbb{W}$ attain the same values $X_T, L_T^r(X)$ then they have the same distributions conditioned on $X_T, L_T^r(X)$. Furthermore the law of $X_T, L_T^r(X)$ of measure $\mathbb{S}$ is given by functions g and g_0. It relates to a similar construction in EA1 using h as the density of biased Brownian motion at time T. Local time at time T and level r has a probability point mass at 0 thus the joint distribution of $X_T, L_T^r(X)$ is of unusual form. Functions g and g_0 are case specific as they include $A(X_T)$ term and α. It is important to notice that the term $e^{-y\frac{\alpha(r+)-\alpha(r-)}{2}}$ is dominated in the

tails by $e^{-\frac{(y+|z-r|+|r-x|)^2}{2s}}$. Sampling from g and g_0 can be performed by rejection sampling with appropriate case specific candidate distribution on $\mathbb{R} \times (0, \infty)$.

We arrive at an algorithm which covers both cases: of positive and of negative value of $\alpha(r_+) - \alpha(r_-)$ and requires only the set of assumptions as on page 34.

Alg.3

INPUT:

$\alpha(u) : \mathbb{R} \to \mathbb{R}$ drift function

$\{0 = t^0, t^1, \cdots, t^n, t^{n+1} = T\}$ time instances of interest

$X_0 = x$ initial point

I **1**. Define

(a) $A : \mathbb{R} \to \mathbb{R}$, $A(u) := \int_0^u \alpha(y)dy$

(b) $k = \inf_{u \in \mathbb{R}} \frac{\alpha^2(u)+\alpha'(u)}{2}$

(c) $\phi : \mathbb{R} \to \mathbb{R}$, $\phi(u) := \frac{\alpha^2(u)+\alpha'(u)}{2} - k$

2. Find the upper bound M of ϕ.

II Generate realisation of X and $L^r(X)$ at T:

Sample X_T, L_T^r such that

$P(X_T \in dz, L_T^r \in dy) = g(z, y)dzdy$ for $y > 0$

and

$P(X_T \in dz, L_T^r = 0) = g_0(z)dz$

III Sample an auxiliary planar Poisson process Ψ with intensity 1

(i) Sample $k \sim \text{Poiss}(TM)$

(ii) Sample $U_1, \cdots, U_k$ i.i.d., $U_1 \sim \text{Uni}((0,T) \times (0,M))$

(iii) Sort $\{u_1, \cdots, u_k\}$ in increasing order of their first coordinate τ to get $((\tau_1, \psi_1), \cdots, (\tau_k, \psi_k))$

IV Generate X jointly with $L^r(X)$ conditioned on values at 0 and T:

(a) Sort in increasing order the set $\{0, \tau_1, \cdots, \tau_k, t^1, \cdots, t^n, T\}$ to get $(0, t_1, \cdots, t_{k+n}, T)$

(b) Repeat $k+n$ times

(i) Pick $i \in \{1, \cdots, k+n\}$ such that X_{t_i} has not been sampled yet

(ii) Find $t_L := \max_{\{0,t_1,\cdots,t_{k+n},T\}}\{t_j$ such that X_{t_j} has been already revealed and $t_j < t_i\}$

(iii) Find $t_U := \min_{\{0,t_1,\cdots,t_{k+n},T\}}\{t_j$ such that X_{t_j} has been already revealed and $t_j > t_i\}$

(iv) Sample $(X_{t_i}, L^r_{t_i})$ using the joint Brownian and local time bridge measure $\mathbb{BB}^{x,b_L,l_L,b_U,l_U}_{t_L,t_U}$ [See Section 3.5]

V Use Ψ to perform rejection sampling:

(a) Compute $\phi(X_{\tau_i})$ for $i \in \{1, \cdots, k\}$

(b) If $\phi(X_{\tau_i}) < \psi_{\tau_i}$ for each $i \in \{1, \cdots, k\}$ then return

$$((0,x,0), (t_1, X_{t_1}, L^r_{t_1}(X)), \cdots, (t_{k+n}, X_{t_{k+n}}, L^r_{t_{k+n}}(X)), (T, X_T, L^r_T(X))$$

Otherwise start again at II.

OUTPUT:

$$((0,x,0),(t_1,X_{t_1},L^r_{t_1}(X)),\cdots,(t_{k+n},X_{t_{k+n}},L^r_{t_{k+n}}(X)),(T,X_T,L^r_T(X))$$

Alg.3 shares with Alg.1 and Alg.2 a property that additional values of X and $L^r(X)$ can be easily obtained after returning output as above. It requires only simulation as in Section 3.5. On the other hand Alg.3 is more widely applicable. However if Assumption 3.2 is satisfied it is easier to sample using Alg.1 or Alg.2.

.

Part II

Curvature Metropolis-adjusted Langevin algorithms

Chapter 5

Introduction

Metropolis-Hastings algorithms have been hugely successful in exploring posterior distributions for Bayesian analysis (see, e.g. [RC04]). Most of the commonly implemented methods such as the random walk Metropolis algorithm and the Langevin-diffusion motivated MALA algorithm, are simple to implement and use constant proposal distribution variance. Moreover, the only user-determined algorithm parameter, the proposal variance, is a well-studied problem (see [RR01] and references therein) and is now routinely calibrated within adaptive MCMC strategies (see e.g. [RR09, HST01, RR07]).

While the strategy of constant proposal variance is simple and often effective, it is common that proposal variances appropriate for some parts of the state space are not suitable for other parts of the space, and it is natural in this case to consider alternatives in which the proposal variance is allowed to vary with the state. Within the context of Langevin algorithms and Hamiltonian MCMC, Mark Girolami and Ben Calderhead in [GC11] introduced an elegant, flexible and powerful framework for this problem by introducing a metric on the parameter space through which state-dependent scaling can be naturally defined.

Our work here is inspired by [GC11], and offers theoretical results to underpin

the state-dependent approach. Our algorithm will be similar to, though will differ from the Manifold MALA algorithm introduced in [GC11]. While Manifold MALA requires the construction of a metric tensor on the state space, for instance constructed from the Fisher's Information matrix for the experiment and model inducing the posterior distribution, our method is somewhat direct. We do not construct a metric, instead we consider the state-dependent scaling problem directly. This has the advantage of being (at least in principle) a more generally applicable approach.

We shall term our method Curvature MALA, and the main contribution to our work will be results that demonstrate the geometric ergodicity for Curvature MALA for light tailed distributions in one dimension. Our calculations concentrate on the case where target densities can be written in the form $\exp\{-|x|^\beta\}$ at least up to proportionality and for x in the tails of the distribution.

The proofs of our results, even in these simple cases, are quite involved. However we do believe they are highly constructive, giving a clear qualitative picture of the behaviour of the algorithm which can readily be transported to other examples according to their target distribution tail behaviour.

Chapter 6

Rationale and development

We take a step back and start with a discussion of generic ways to improve and speed up Metropolis-Hastings based MCMC algorithms. Our Curvature MALA algorithm will arise as a natural improvement of the MALA algorithm by considering a non-constant local Euler discretization of the underlying Langevin diffusion. In what follows, let π be the target distribution on a general state space $\mathcal{X}$ and write P and Q for Markov transition kernels on $\mathcal{X}$. We refer to [RR04] for introduction to Markov chains and general state space MCMC algorithms.

The Metropolis-Hastings algorithm [Has70] is an extremely general tool that given any irreducible Markov transition kernel Q with transition densities $q(x, y)$ constructs a reversible and ergodic Markov chain P with stationary distribution π by thinning the dynamics of Q with the accept-reject ratio of the form

$$\alpha(x, y) \quad = \quad \min\{1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\}. \tag{6.1}$$

Since this happens under very mild conditions (see e.g. [Tie98]), the algorithm is a victim of its own generality in the sense that poor choice of the proposal kernel Q results in slow convergence and the optimal choice of tractable Q is not known. Guidance of how to optimize the choice of Q is available only within very specific

classes of kernels.

The first fundamental example is the choice of variance for the Gaussian random walk Metropolis proposals on $\mathcal{X} = \mathbb{R}^d$, i.e. for the case

$$Q(x, \cdot) \quad = \quad N(x, \Sigma). \tag{6.2}$$

The efficiency of the algorithm depends on Σ and has been studied formally via diffusion limits in high dimensions in [RGG97, RR01]. Advice from these results is routinely used within adaptive MCMC algorithms [RR09, HST01, RR07].

The next natural question is whether one can also improve on the mean of the normal proposal e.g. by biasing the algorithm towards a region of higher stationary density. This idea has been formally addressed in [RT96a] and then [ST99a, ST99b], by employing Langevin dynamics, i.e. continuous time diffusion process of the form

$$dX_t \quad = \quad \frac{1}{2}\nabla \log(\pi(X_t))dt + dB_t \tag{6.3}$$

designed to have π as stationary measure (see also [RDF78] for related work in the applied context). An Euler discretization of (6.3) with step h is then used as a Metropolis-Hastings proposal resulting in

$$Q(x, \cdot) \quad = \quad N\Big(x + \frac{h}{2}\nabla \log \pi(x),\ h\Big), \tag{6.4}$$

and the algorithm is termed MALA (Metropolis adjusted Langevin algorithm). The optimal choice of h has been considered in [RR98, RR01], again by analysing the performance of the algorithm as the dimension of the state space $\mathcal{X}$ grows to infinity. An adaptive MALA algorithm utilizing these results has been proposed in [Atc06] and [MR11].

Here we employ a different rationale behind choosing h, one that allows for the discretization to be state-dependent and does not rest on high dimensional limits.

In particular, since the MALA proposal (6.4) moves towards the high probability region by following $\nabla \log \pi(x)$, the first observation is that if $\nabla \log \pi(x)$ is not stable around x, it should not be 'trusted' and h should be chosen small. To measure the stability of $\nabla \log \pi(x)$ we compute the Hessian of $\log \pi(x)$. If the absolute value of an eigenvalue of $H(\log \pi(x))$ is large, $\nabla \log \pi(x)$ will change rapidly when moving along the respective eigenvector and the direction should be penalised appropriately in h. We thus make use of eigenvectors and absolute values of the eigenvalues of $H(\log \pi(x))$ to construct h as a transformation to be applied to $\nabla \log \pi(x)$, rather than a number. To this end write first the eigenvalue decomposition

$$H(\log \pi(x)) \;=\; V \Lambda V^T, \tag{6.5}$$

where Λ is the diagonal matrix of eigenvalues. Now let Λ_{abs} be the diagonal matrix of absolute values of eigenvalues and define the state dependent matrix h as

$$h(x) \;:=\; V \Lambda_{\text{abs}}^{-1} V^T. \tag{6.6}$$

This curvature based local discretization $h(x)$ results in the new proposal

$$Q(x, \cdot) \;=\; N\Big(x + \frac{h(x)}{2} \nabla \log \pi(x),\; h(x)\Big). \tag{6.7}$$

It can be viewed as discretization with step size 1 of

$$dX_t \;=\; \frac{\sigma^2(X_t)}{2} \nabla \log \pi(X_t) dt + \sigma(X_t) dB_t, \tag{6.8}$$

where $\sigma^2(x) := h(x)$ given by (6.6) and where we use the convention $\sigma(x)\sigma'(x) = \sigma^2(x)$. The diffusion (6.8) does not preserve π as its stationary distribution, however this can be fixed by introducing a slight correction. Under mild regularity conditions (see [Ken78], [XSL+13]) any choice of $\sigma(\cdot)$ in

$$dX_t \;=\; \Big(\frac{\sigma^2(X_t)}{2} \nabla \log \pi(X_t) + c(X_t)\Big) dt + \sigma(X_t) dB_t, \tag{6.9}$$

where the correction term

$$c(x)_i \;:=\; \frac{1}{2}\sum_j \frac{\partial}{\partial x_j}(\sigma^2(x))_{ij}, \tag{6.10}$$

yields a diffusion with stationary distribution π. Thus instead of discretising (6.8) with stepsize 1, one can discretise the corrected diffusion (6.10) with arbitrary stepsize h arriving at proposals

$$Q(x,\cdot) \;=\; N\Big(x + h\mu(x),\; h\sigma^2(x)\Big), \qquad \text{where} \tag{6.11}$$

$$\mu(x) \;:=\; \frac{\sigma^2(x)}{2}\nabla \log \pi(x) + c(x), \tag{6.12}$$

and $\sigma^2(x) = h(x)$ defined as before by (6.6). This yields the Curvature MALA algorithm (CMALA), which dynamics for fixed discretization step h is given by

Algorithm 1 (Curvature MALA).

Given $X_n = x$,

1. *Compute* $\sigma^2(x)$ *from* (6.5) *and* (6.6) *and* $\mu(x)$ *from* (6.12),
2. *Draw* $Y_n \sim Q(x,\cdot)$ *of* (6.11),
3. *Put* $X_{n+1} = Y_n$ *or* $X_{n+1} = X_n$ *according to the Metropolis-Hasting accept-reject ratio.*

In fact, as it happens for the benchmark family of target distributions with tail behaviour $\propto \exp\{-|x|^\beta\}$, the correction term $c(x)$ defined in (6.10) is often negligible, at least in the tails, compared to the main term $\frac{\sigma^2(X_t)}{2}\nabla \log \pi(X_t)$, called the *net drift* [Ken78]. Furthermore under the assumption that $c(x)$ is negligible significant implementation speed-ups can be achieved by dropping the $c(x)$ term which motivates the simplified Net Curvature MALA (netCMALA) whose proposal process follows the discretisation of the initial diffusion (6.8), i.e. where (6.11)

and (6.12) are replaced by

$$Q_{\text{net}}(x,\cdot) = N\Big(x + h\mu_{\text{net}}(x),\ h\sigma^2(x)\Big), \qquad \text{where} \tag{6.13}$$

$$\mu_{\text{net}}(x) := \frac{\sigma^2(x)}{2}\nabla \log \pi(x), \tag{6.14}$$

Algorithm 2 (Net Curvature MALA). *Given* $X_n = x$,

1. *Compute* $\sigma^2(x)$ *from* (6.5) *and* (6.6) *and* $\mu_{net}(x)$ *from* (6.14)
2. *Draw* $Y_n \sim Q_{net}(x,\cdot)$ *of* (6.13),
3. *Put* $X_{n+1} = Y_n$ *or* $X_{n+1} = X_n$ *according to the Metropolis-Hasting accept-reject ratio.*

We shall prove geometric convergence of CMALA and netCMALA under some regularity conditions and for h small enough. It is rarely known how small is small enough, but the problem can be alleviated by making h random. In other words h is sampled from some fixed probability measure ν on $(0,\infty)$. It is of practical importance to chose ν such that $\nu((0,\epsilon)) > 0$ for all $\epsilon > 0$. Then there is always positive probability that h will be sufficiently small. As we see in Chapter 7 this change does not weaken the results.

Another technical burden is that (net)CMALA is not defined if one or more of the eigenvalues of Λ are 0. We interpret that in this case the variance of the proposal goes to infinity and the proposed move is always rejected and the algorithm gets stuck in one state. To overcome this difficulty and let the algorithm explore again entire density function we suggest to add a standard random walk Metropolis (RWM) move whenever it happens. Thus finally we arrive at the Randomized (Net) Curvature MALA which amounts to the following steps:

Algorithm 3 (R(net)CMALA). *Let $p \in (0,1)$ be fixed and suppose ν is a probability measure on $(0,\infty)$.*
Given $X_n = x$,

1. *Draw $U \sim \mathrm{Bernoulli(p)}$,*
2. *If $U = 1$ and 0 is not an eigenvalue of Λ,*
   - *draw $h \sim \nu$,*
   - *find X_{n+1} using the (net)CMALA kernel with discretization h.*
3. *If $U = 1$ and 0 is an eigenvalue of Λ,*
   - *set $X_{n+1} = x$.*
4. *If $U = 0$,*
   - *find X_{n+1} using the RWM kernel with proposal distribution $Q(x,\cdot) = \mathbb{N}(x, I)$.*

If ν is a continuous probability distribution then one needs to check measurability of the resulting transition operator to ensure it is well-defined (see e.g. [Num04]). If ν is discrete these conditions are always met. In the rest of the paper we assume the kernel is well-defined.
As we have already mentioned, what makes the defined algorithms special is the choice of a state-dependent scaling. Exploration of the target density depends on the direction and size of moves of algorithms. Gradient of $\log(\pi(x))$ indicates the direction of a fastest ascend of the target density function at point x. When the rate of change of gradient is slow, what occurs for the tail behaviour we are interested in, it is important to direct the moves towards regions of high density and make big moves. If we have a look at (6.6), the new eigenvalues, where each of them is the inverse of the absolute value of an eigenvalue of Hessian of

$\log(\pi(x))$, are used to balance the rate of change of the gradient and they promote big moves when the density is relatively flat. In R(net)CMALA we suggest to use RWM move when any of the eigenvalues is equal to zero. R(net)CMALA could be further modified in a way that we choose to perform such step in case that any of the eigenvalues is in some fixed small interval around zero. It would allow us to avoid proposals with too big variance. Addition of an RWM move can also help in case of multimodality as discussed algorithms not only benefit from using local geometry but unfortunately also suffer from it. The proposal distribution is tailored to local geometry of target density and it does not address global structure. Locally optimal moves often do not favour visits to multiple modes. In the regions where target density gets steeper and the change of gradient is fast the choice of scaling as described encourages smaller moves. It allows for better control of the algorithm to stay in the region of high density. Our intuition is that the suggested algorithms will not perform very well in case of multimodality, heterogeneity and with heavy tails.

Connection to Manifold MALA of [GC11]

The diffusion (6.9) with stationary distribution π may appear very flexible and powerful as a tool for designing Metropolis-Hastings proposals, however sensible choice of non-constant $\sigma(x)$ is nonobvious (see [ST99a, ST99b, RS02] for related work in this direction). Within the context of Bayesian posterior computation, [GC11] introduced a general framework where $\sigma(x)$ results from a metric tensor defining a Riemannian manifold on the parameter space. The choice of the metric tensor in [GC11] is again not unique and can be problematic due to positive definiteness requirement. The performance of the algorithm will also depend heavily on this choice.

Although the rationale we develop here for Curvature MALA does not result from considering a metric on probability distributions, and the algorithm is constructed more directly, there are strong ties with the algorithms suggested in [GC11]. Under the assumptions (R1), (R2), (R3) of following Chapter 7 the Curvature MALA is equivalent to the Manifold MALA with the metric tensor taken as the observed Fisher information matrix plus the negative Hessian of the log-prior. Note that the algorithms in [GC11] are considered in a Bayesian setting and they aim to sample from a posterior distribution of parameter $\theta \in \Theta$ which corresponds to our $x \in X$. Furthermore recall that the observed Fisher information matrix is the negative of the Hessian matrix of the log-likelihood. The above statement can be further compared with third paragraph on 129 of citegirolami2011riemann, Similarly, under assumptions (R1), (R2*), (R3) the Net Curvature MALA becomes the Simplified Manifold MALA defined on page 130 between equations (10) and (11). We would like to underline that our work is one of the first attempts to provide convergence analysis of methods related to [GC11] where lots of fruitful ideas where presented with illustrative examples. In the following Chapter 7 we discuss results on geometric ergodicity of (R)(net)CMALA.

Chapter 7

Results for Curvature MALA and its modifications.

We are interested here in properties of the algorithms introduced in Chapter 6. In particular, we investigate under what conditions the algorithms are geometrically ergodic. We understand geometric ergodicity as defined in [RR04].

Definition 7.1. *A Markov chain with stationary distribution $\pi(\cdot)$ is geometrically ergodic if*

$$||P^n(x,\cdot) - \pi(\cdot)||_{TV} \leq M(x)\rho^n \quad n = 1, 2, 3, \ldots \tag{7.1}$$

for some $\rho < 1$, where $M(x) < \infty$ for $\pi - a.e.\ x \in X$.

Recall that the *total variation distance* is defined for every two probability measures μ and ν on the same probability space as $||\mu - \nu||_{TV} := \sup_{B\in\mathcal{B}} |\mu(B) - \nu(B)|$. Further, we say that a *$\sqrt{n}$-central limit theorem* holds for square-integrable with respect to π function f and $(X_n)_{n\geq 0}$ if there exists $\sigma_f \in [0, \infty)$ such that $\frac{S_n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma_f^2)$ as $n \to \infty$, where $S_n := \sum_{i=0}^{n-1} \bar{f}(X_i)$ and $\bar{f}$ is the centered version of f, namely $\bar{f} = f - \int_X f(x)\pi(dx)$.

Geometric ergodicity is of crucial importance in the context of MCMC algorithms because for reversible Markov chains this property implies a central limit theorem for functions that are square-integrable with respect to π.

It is highly desirable for the computations to be asymptotically validated by a CLT and asymptotic confidence intervals rather than by a law of large numbers or merely by convergence in total variation. A CLT for Markov chains is guaranteed under varied conditions, see [J+04],[RR04] for reviews. Curvature MALA is reversible thus the following result proved in [RR97] applies

Theorem 7.2. *If a Markov chain $(X_n)_{n\geq 0}$ with stationary distribution π is geometrically ergodic and reversible, then a $\sqrt{n}-$CLT holds for $(X_n)_{n\geq 0}$ and f whenever $\pi(f^2) < \infty$.*

In fact there is even stronger result in [RR08], using variance bounding chains, which implies that if a reversible Markov chain is not geometrically ergodic then for some L^2 functions CLT does not hold.

Now we proceed to discussion on earlier defined algorithms applied to target distribution π defined on $\mathbb{R}$. Proposals in CMALA and netCMALA then follow normal distributions $Q_C(x,\cdot) = N(\mu_x^C, \sigma_x^2)$ and $Q_{net}(x,\cdot) = N(\mu_x^{net}, \sigma_x^2)$ respectively with parameters

$$\begin{aligned} \mu_x^C &= x + h\Big[\frac{1}{2}\sigma^2(x)\frac{\pi'(x)}{\pi(x)} + \sigma(x)\sigma'(x)\Big], \\ \mu_x^{net} &= x + h\frac{1}{2}\sigma^2(x)\frac{\pi'(x)}{\pi(x)}, \\ \sigma_x^2 &= h\sigma^2(x), \\ \text{where}\quad \sigma(x) &= \Big|\frac{\partial^2}{\partial x^2}\log\pi(x)\Big|^{-\frac{1}{2}} = \Big|\frac{\pi''(x)\pi(x) - (\pi'(x))^2}{\pi^2(x)}\Big|^{-\frac{1}{2}}. \end{aligned}$$

In particular, we investigate the convergence rate in total variation norm of the algorithms applied to a benchmark family of distributions π with $\pi(x) \propto \exp\{-|x|^\beta\}$

in the tails. More precisely, we allow for different β's in the left and right tail, i.e. for some $M > 0$, $k_+ > 0$ and $k_- > 0$ we assume

(R0)
$$\begin{aligned} \pi(x) &= k_+ \exp\{-|x|^{\beta_+}\} \qquad \text{for all} \quad x > M, \\ \pi(x) &= k_- \exp\{-|x|^{\beta_-}\} \qquad \text{for all} \quad x < -M. \end{aligned}$$

Then for sufficiently large x the parameters of the proposal distribution become

$$\mu_x^C = x[1 - \frac{h}{2|\beta_x - 1|}(1 + \frac{\beta_x - 2}{\beta_x}|x|^{-\beta_x})], \tag{7.2}$$

$$\mu_x^{net} = x[1 - \frac{h}{2|\beta_x - 1|}] \tag{7.3}$$

$$\sigma_x^2 = h\frac{|x|^{2-\beta_x}}{|\beta_x - 1|\beta_x}, \tag{7.4}$$

where β_x is equal to β_+ or β_- respectively to the sign of x. Let $l(x) := \frac{\partial^2}{\partial x^2} \log \pi(x)$ and λ denotes Lebesgue measure. In order to implement the algorithms of Section 2 we need some regularity conditions:

(R1) $\pi(x) > 0$ for all $x \in \mathbb{R}$

(R2) $\pi \in C^3$

(R3) $l(x) \neq 0$ for all $x \in \mathbb{R}$

In some cases, which are specified in the following theorems, (R2) and (R3) can be replaced by the following weaker assumptions

(R2*) $\pi \in C^2$

(R3*) $l(x) \neq 0$ for $x \in \mathbb{R} \setminus D$ where D is a bounded subset of $\mathbb{R}$ with $\lambda(D) = 0$

Now we present our key results on convergence of (net)CMALA. Proofs of all Theorems are deferred until next Chapter.

Theorem 7.3.

(a) *Suppose (R0), (R1), (R2*) and (R3) hold. If also the following conditions are satisfied*

(C1.1) $\beta_+ > 1$ *and* $\beta_- > 1$

(C2.1) $h > 0$ *is sufficiently small*

then netCMALA is well-defined and geometrically ergodic.

(b) *If all conditions in (a) are satisfied and in addition (R2) holds then also CMALA is well-defined and geometrically ergodic.*

Theorem 7.4.

(a) *Suppose (R0), (R1), (R2*) and (R3) hold. If also the following condition is satisfied*

(C1.2) $\beta_+ \neq \beta_-$, $\min\{\beta_+, \beta_-\} \in (0, 1)$

then netCMALA is well-defined but not geometrically ergodic.

(b) *If all conditions in (a) are satisfied and in addition (R2) holds then also CMALA is well-defined but not geometrically ergodic.*

The difficulty in applying Theorem 7.3 is that it assumes strong regularity conditions on π and the knowledge of how small h needs to be. These requirements can be alleviated by using R(net)CMALA, as demonstrated in Theorem 7.5 below.

Theorem 7.5.

(a) *Suppose (R0), (R1), (R2*) and (R3*) hold. If also the following conditions are satisfied*

(C1.1) $\beta_+ > 1$ *and* $\beta_- > 1$

(C2.2) *ν is discrete and satisfies $\nu((0,\varepsilon)) > 0$ for all $\varepsilon > 0$*

then RnetCMALA is well-defined and geometrically ergodic.

(b) *If all conditions in (a) are satisfied and in addition (R2) holds then also RCMALA is well-defined and geometrically ergodic.*

One could hope that lack of geometric ergodicity, as established in Theorem 7.4, could be addressed by randomizing h. However this is not the case. More precisely, we prove the following Theorem 7.6.

Theorem 7.6.

(a) *Suppose (R0), (R1), (R2*) and (R3*) hold. If also the following conditions hold*

(C1.2) $\beta_+ \neq \beta_-$, $\min\{\beta_+, \beta_-\} \in (0,1)$

(C2.3) *ν is a discrete probability measure*

then RnetCMALA is well-defined but not geometrically ergodic.

(b) *If all conditions in (a) are satisfied and in addition (R2) holds then also RCMALA is well-defined but not geometrically ergodic.*

In case of distributions with symmetric heavy tails net(CMALA) in general is not geometrically ergodic. However for h in some small interval it is geometrically ergodic. It is interesting from theoretical point of view but we believe it is not of practical importance.

Proposition 7.7.

(a) *Suppose (R0),(R1), (R2*) and (R3) hold. If the following condition hold*

(C1.3) $\beta \in (0,1)$

then netCMALA is well-defined and there exists $\eta \in (2, 4)$ *such that if*

(C2.4) $h \in (\eta|\beta - 1|, 4|\beta - 1|)$

then netCMALA is geometrically ergodic and if

(C2.5) $h \in (0, \eta|\beta - 1|) \cup (4|\beta - 1|, \infty)$

then netCMALA is not geometrically ergodic.

(b) *If (R2*) is replaced by (R2), the statement in part (a) is true also for CMALA.*

Chapter 8

Proofs

Before we start presentation of the proofs of the results of Chapter 7 we give a brief overview of the concepts that are important for the proofs. Then we concentrate on proving geometric ergodicity, as defined in Definition 7.1, of (net)CMALA for a benchmark family of distributions with light tails (Theorem 7.3) and a special case within a family of symmetrically heavy-tailed distributions (Proposition 7.7). Next we show lack of geometric ergodicity in case of heavy-tailed distributions (Theorem 7.4 and Proposition 7.7). Then we proceed to the proof of Theorem 7.5 for algorithms with randomized h. Throughout all proofs we assume without loss of generality that $\beta_- \geq \beta_+$.

Here we briefly describe those concepts which are used in the proof of geometric ergodicity. For further details, see e.g. [MT93] or [RR04]. Our primary interest lies in probability density function π defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We call $\mathbb{R}$ a *state space* and denote it by X. We refer to its elements as *states* and denote them usually by x's or y's. The algorithms constructed in the previous chapters give rise to Markov chains on X. We use standard notation for transition probability kernel with subscripts indicating the algorithm in consideration. Namely, $P_{CMALA}(x, A)$ with $x \in X$ and $A \subset X$ denotes probability of moving from state x to set A in

one move of CMALA algorithm. Further recall that a set $C \in \mathcal{B}(X)$ is called a *small set* if there exist a positive integer n_0, $\varepsilon > 0$ and a probability measure $\nu(\cdot)$ on X such that the following *minorisation condition* holds: $P^{n_0}(x, A) \geq \varepsilon\nu(A)$ for $x \in C$ and $A \in \mathcal{B}(X)$. We say that a Markov chain satisfies a *drift condition* if there exist constants $0 < \lambda < 1$ and $b < \infty$, a measurable set C and a function $V : X \to [1, \infty]$, such that

$$\int_X P(x, dy)V(y) \quad \leq \quad \lambda V(x) + b\mathbb{1}_C(x), \quad x \in X. \tag{8.1}$$

The following Theorem 8.1 of Meyn and Tweedie [MT93] is widely exploited in proofs of geometric ergodicity of Markov chains; we use the version as in [RR04].

Theorem 8.1. *Consider a ϕ-irreducible, aperiodic Markov chain with stationary distribution $\pi(\cdot)$. Suppose the minorisation condition is satisfied for some $C \subset X$ and $\varepsilon > 0$ and probability measure $\nu(\cdot)$. Suppose further that the drift condition is satisfied for some constants $0 < \lambda < 1$ and $b < \infty$, and a function $V : X \to [1, \infty]$ with $V(x) < \infty$ for at least one (and hence for π-a.e.) $x \in X$. Then the chain is geometrically ergodic.*

Next we describe a link between the spectral gap of Markov operator and its rate of convergence which we will use later to show that the results also hold with randomized h. Consider a Hilbert space $L^2(\pi)$ of functions $f : \mathbb{R} \to \mathbb{C}$ whose $L_2(\pi)$ norm, induced by the following inner product, is finite; $\langle f, g\rangle := \int_{\mathbb{R}} f\ \overline{g}\ d\pi$ where $\overline{g}$ denotes the complex conjugate of g. We say that σ_P is an $L^2(\pi)$-spectrum of a Markov operator P if it contains all $\lambda \in \mathbb{C}$ for which the inverse of $(\lambda I - P)$ does not exist as a bounded linear operator on $L^2(\pi)$. Note that every reversible Markov chain possesses a real spectrum. Let P restricted to $\mathbf{1}^\perp$ be denoted by $P|_{\mathbf{1}^\perp}$. Define $\delta_P^{gap} := 1 - \sup\{|\lambda| : \lambda \in \sigma_{P|_{\mathbf{1}^\perp}}\}$ and $\delta_P := 1 - \sup\{\lambda : \lambda \in \sigma_{P|_{\mathbf{1}^\perp}}\}$. Recall that transition kernel P of ϕ-irreducible, aperiodic Markov chain admits an $L^2(\pi)$-spectral gap if a spectrum of $P|_{\mathbf{1}^\perp}$ is bounded away from the unit circle.

Then $\delta_P^{gap} > 0$ is called the spectral gap of P. Furthermore, note that

Proposition 8.2 (Proposition 1.2 of [KM09]). *A reversible, ψ-irreducible and aperiodic Markov chain $(X_n)_{n\in\mathbb{N}}$ is geometrically ergodic if and only if P admits a spectral gap in $L^2(\pi)$.*

Now let $k_P(A)$ denote the rate of probability flow from a set A to A^c, normalized by the product of $\pi(A)$ and $\pi(A^c)$.

$$k_P(A) := \frac{\int_A \pi(dx)P(x, A^C)}{\pi(A)\pi(A^c)}.$$

Suppose $k_P := \inf_{A\in\mathcal{B}(\mathbb{R}):\, 0<\pi(A)<1} k_P(A)$ which is an analogue of Cheeger's isoperimetric constant. Then the Theorem 2.1 of [LS88] gives the following bounds

$$\frac{k_P^2}{8} \leq \delta_P \leq k_P. \tag{8.2}$$

Now we proceed to the proofs of our main results.

Proof of Theorem 7.3. Suppose that suitable regularity conditions R0, R1, R2(R2*) and R3 hold. We begin with discussion on the acceptance probability $\alpha(x, y)$, defined in (6.1), in (net)CMALA under assumption that x and y are sufficiently far in the tails. In particular, we assume that $\pi(x) = k_+ \exp\{-|x|^{\beta_+}\}$ and

$$\pi(y) = k_+ \exp\{-|y|^{\beta_+}\} \quad \text{or} \quad \pi(y) = k_- \exp\{-|y|^{\beta_-}\}.$$

Let then $\beta_y = \beta_+$ or $\beta_y = \beta_-$ respectively, similarly for k_y. Saying more intuitively, x is in the right tail while y is either also in the right one or it is in the left tail. Similar computations for $\alpha(x, y)$ to the following ones can be performed when $x < -M$ as in (R0). We use superscripts ^{C} and ^{net} to refer to CMALA and netCMALA respectively.

In order to aggregate variables in a useful manner for the remainder of the proof we introduce variables $\mathsf{a}, \mathsf{a}_y, \mathsf{b}, \mathsf{c}$ and d satisfying

$$\begin{aligned}
\mu_x^{net} &= (1-\mathsf{a})x, & \mathsf{a} &= \frac{h}{2|\beta_+ - 1|},\\
\mu_y^{net} &= (1-\mathsf{a}_y)y, & \mathsf{a}_y &= \frac{h}{2|\beta_y - 1|},\\
\mu_x^{C} &= (1-\mathsf{b})x, & \mathsf{b} &= \mathsf{a}(1+\frac{\beta_+ - 2}{\beta_+}|x|^{-\beta_+})\\
\mu_y^{C} &= (1-\mathsf{c})y, & \mathsf{c} &= \mathsf{a}_y(1+\frac{\beta_y - 2}{\beta_y}|y|^{-\beta_y})\\
y &= (1-\mathsf{d})x.
\end{aligned} \tag{8.3}$$

Then

$$\begin{aligned}
\alpha^C(x,y) &= \min\left\{1, \frac{k_y e^{-|y|^{\beta_y}} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(x-\mu_y^C)^2}{2\sigma_y^2}}}{k_+ e^{-|x|^{\beta_+}} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(y-\mu_x^C)^2}{2\sigma_x^2}}}\right\}\\
&= \min\Bigg\{1, \frac{k_y}{k_+}|x|^{\frac{\beta_y-\beta_+}{2}}|1-\mathsf{d}|^{\frac{\beta_y}{2}-1}(|\beta_y-1|\beta_y)^{\frac{1}{2}}(|\beta_+-1|\beta_+)^{-\frac{1}{2}}\\
&\quad \times e^{|x|^{\beta_+}\cdot(1+\frac{\beta_+(\mathsf{b}-\mathsf{d})^2}{4\mathsf{a}})-|x|^{\beta_y}|1-\mathsf{d}|^{\beta_y}(1+\frac{\beta_y(1-(1-\mathsf{c})(1-\mathsf{d}))^2}{4\mathsf{a}_y|1-\mathsf{d}|^2})}\Bigg\}.
\end{aligned} \tag{8.4}$$

We deal separately with the cases where $\beta_y = \beta_+$ and $\beta_y = \beta_-$. If $\beta_y = \beta_+$, which in particular occurs when $y > M$ and when $\beta_- = \beta_+$ and y is in either of the tails, then (8.4) reduces to

$$\alpha^C(x,y) = \min\left\{1, \frac{k_y}{k_+}|1-\mathsf{d}|^{\frac{\beta_+}{2}-1} e^{|x|^{\beta_+}\cdot[1-|1-\mathsf{d}|^{\beta_+}+\frac{\beta_+}{4\mathsf{a}}[(\mathsf{b}-\mathsf{d})^2-(\mathsf{c}+\mathsf{d}-\mathsf{c}\mathsf{d})^2|1-\mathsf{d}|^{\beta_+-2}]]}\right\}. \tag{8.5}$$

Let $g : (\mathbb{R}\setminus\{0\}) \times \mathbb{R}^2 \times (\mathbb{R}\setminus\{1\}) \to \mathbb{R}$ be defined as follows

$$g(a,b,c,d) = 1 - |1-d|^{\beta_+} + \frac{\beta_+}{4a}[(b-d)^2 - (c+d-cd)^2|1-d|^{\beta_+-2}].$$

Further, we write (8.5) as

$$\alpha^C(x,y) = \min\left\{1, \frac{k_y}{k_+}|1-\mathsf{d}|^{\frac{\beta_+}{2}-1} e^{|x|^{\beta_+} g(\mathsf{a},\mathsf{b},\mathsf{c},\mathsf{d})}\right\}. \tag{8.6}$$

Note that a is constant throughout whole exposition here. So far we have considered $\alpha(x,y)$ where we treated x and y as fixed. All new variables in (8.3) were

determined by this pair of x and y. However now we allow for x to change and we treat y as function of x given by the bottom equation in (8.3). Thus let for now d (d $\neq 1$) be constant. We treat other new variables in (8.3) as functions of x and y (so also implicitly of x). We will investigate now changes in these variables when x goes to ∞. Then remembering that here $\beta_y = \beta_+$

$$\mathsf{b} = \mathsf{a}(1 + \frac{\beta_+ - 2}{\beta_+}|x|^{-\beta_+}) \xrightarrow{x\to\infty} \mathsf{a} \quad \text{and} \quad \mathsf{c} = \mathsf{a}_y(1 + \frac{\beta_y - 2}{\beta_y}|y|^{-\beta_y}) \xrightarrow{x\to\infty} \mathsf{a}. \tag{8.7}$$

Suppose a $\neq 1$. Motivated by (8.7) we set d = a. Then $\lim_{x\to\infty} \frac{\mu_x^C - y}{x} = 0$. Further it suggests to consider smooth $f : (\mathbb{R} \setminus \{1\}) \to \mathbb{R}$ satisfying $f(a) = g(a, a, a, a)$ for $a \in \mathbb{R} \setminus \{0, 1\}$, and compute its derivatives. Observe that

$$\begin{aligned}
f(a) &= 1 - |1-a|^{\beta_+} - \frac{\beta_+}{4}a(2-a)^2|1-a|^{\beta_+-2} \\
f'(a) &= \operatorname{sgn}(1-a)\frac{\beta_+}{4}a|1-a|^{\beta_+-3}\,[\,(1+\beta_+)a^2 + (1-4\beta_+)a + 4\beta_+ - 4] \\
f(0) &= 0 \\
f'(0) &= 0 \\
f''(0) &= \beta_+(\beta_+ - 1).
\end{aligned} \tag{8.8}$$

Under assumption that $\beta_y = \beta_+$, we also note that

$$\alpha^{net}(x, y) = \min\{1, \frac{k_y}{k_+}|1 - \mathsf{d}|^{\frac{\beta_+}{2}-1} e^{|x|^{\beta_+} g(\mathsf{a},\mathsf{a},\mathsf{a},\mathsf{d})}\}.$$

Crude intuition behind the two following Lemmas is that when current state x is further and further in the tail then mean of proposal distribution μ_x is relatively closer and closer to $(1 - \mathsf{a})x$ and the region with high proposal density, again relatively, stays small around μ_x. Furthermore proposals falling into this region are always accepted. Now we change the framework. We allow x and y to change independently and we treat a_y, b, c and d, as functions of a, x and y. All relations in (8.3) still hold.

Lemma 8.3. *Suppose the assumptions of Theorem 7.3 hold. Then there exists $\varepsilon > 0$ such that for all sufficiently large x*

$$\inf_{y\in A_x} \alpha^C(x, y) = 1 \quad \textit{and} \quad \inf_{y\in A_x} \alpha^{net}(x, y) = 1$$

where $A_x = [(1 - \mathsf{a} - \frac{\varepsilon}{2})x, (1 - \mathsf{a} + \frac{\varepsilon}{2})x]$.

Proof. Referring to (8.8) $f(0) = f'(0) = 0$ and $f''(0) > 0$ thus there exists $0 < \delta < 1$ such that $f(a) > 0$ for all $a \in (0, \delta)$. For all sufficiently small $h > 0$ we have $0 < \mathsf{a} < \delta < 1$ and hence $f(\mathsf{a}) > 0$. Recall that $g(\mathsf{a}, \mathsf{a}, \mathsf{a}, \mathsf{a}) = f(\mathsf{a})$ provided $\mathsf{a} \neq 1$ and $\mathsf{a} \neq 0$. Since g is continuous on its domain there exists $0 < \varepsilon < 1$ such that neither 0 nor 1 is in $[\mathsf{a} - \varepsilon, \mathsf{a} + \varepsilon]$ and for (a, b, c, d) in $[\mathsf{a} - \varepsilon, \mathsf{a} + \varepsilon]^4$ it holds that $g(a, b, c, d) \geq \frac{f(\mathsf{a})}{2} > 0$. Note that we can choose h small enough so that, for sufficiently large x, a is arbitrarily close to zero and $(1 - \mathsf{a} - \frac{\varepsilon}{2})x > M_+$. Furthermore, if $y \in A_x = [(1 - \mathsf{a} - \frac{\varepsilon}{2})x, (1 - \mathsf{a} + \frac{\varepsilon}{2})x]$ then $\mathsf{d} \in [\mathsf{a} - \frac{\varepsilon}{2}, \mathsf{a} + \frac{\varepsilon}{2}]$ and, for sufficiently large x, $\mathsf{c} = \mathsf{a}(1 + \frac{\beta_+ - 2}{\beta_+}|y|^{-\beta_+}) \in [\mathsf{a} - \varepsilon, \mathsf{a} + \varepsilon]$ and $\mathsf{b} = \mathsf{a}(1 + \frac{\beta_+ - 2}{\beta_+}|x|^{-\beta_+}) \in [\mathsf{a} - \varepsilon, \mathsf{a} + \varepsilon]$. Hence $(\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}) \in [\mathsf{a} - \varepsilon, \mathsf{a} + \varepsilon]^4$, $(\mathsf{a}, \mathsf{a}, \mathsf{a}, \mathsf{d}) \in [\mathsf{a} - \varepsilon, \mathsf{a} + \varepsilon]^4$ and so $g(\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}) \geq \frac{f(\mathsf{a})}{2} > 0$ and $g(\mathsf{a}, \mathsf{a}, \mathsf{a}, \mathsf{d}) \geq \frac{f(\mathsf{a})}{2} > 0$. We arrive at

$$\begin{aligned}
\inf_{y \in A_x} \alpha^C(x, y) &= \inf_{y \in A_x} \min\left\{1, \frac{k_y}{k_+}\left|1 - \mathsf{d}\right|^{\frac{\beta_+}{2} - 1} e^{|x|^{\beta_+} g(\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d})}\right\} \\
&\geq \min\left\{1, \inf_{d \in [\mathsf{a} - \varepsilon, \mathsf{a} + \varepsilon]} \left|1 - d\right|^{\frac{\beta_+}{2} - 1} e^{|x|^{\beta_+} \frac{f(\mathsf{a})}{2}}\right\} = 1 \\
\inf_{y \in A_x} \alpha^{net}(x, y) &= \inf_{y \in A_x} \min\left\{1, \frac{k_y}{k_+}\left|1 - \mathsf{d}\right|^{\frac{\beta_+}{2} - 1} e^{|x|^{\beta_+} g(\mathsf{a}, \mathsf{a}, \mathsf{a}, \mathsf{d})}\right\} = 1.
\end{aligned}$$

Lemma 8.4. *Let A_x be constructed as in the Proof of Lemma 8.3. Then $\lim_{x \to \infty} Q_C(x, A_x) = 1$ and $\lim_{x \to \infty} Q_{net}(x, A_x) = 1$.*

Proof. We choose ε as in the Proof of Lemma 8.3. Consider an arbitrary $p \in (0.5, 1)$. Denote by Φ^{-1} the quantile function of $N(0, 1)$. Since $\lim_{x \to \infty} \frac{\mu_x^C}{x} = 1 - \mathsf{a}$ and $\lim_{x \to \infty} \frac{\sigma_x}{x} = 0$, then for sufficiently large x we have $\mu_x^C + \sigma_x \Phi^{-1}(p) \in A_x$ and $\mu_x^C - \sigma_x \Phi^{-1}(p) \in A_x$. Thus $Q_C(x, A_x) \geq 2p - 1$ and this finishes the proof since p was chosen arbitrarily. Analogous computations apply for Q_{net}.

We now use Lemmas 8.3 and 8.4 to check the assumptions of Theorem 8.1 with $V(x)$ chosen to be $|x| + 1$ to confirm geometric ergodicity of CMALA and netCMALA. We present computations only for CMALA as they apply equally to both.

Suppose set A_x is chosen as in the Proof of Lemma 8.3. We have

$$\begin{aligned}\int_{\mathbb{R}} P(x,y)|y|dy &= \underbrace{\int_{A_x} q(x,y)\alpha(x,y)|y|dy}_{=:J_{\mathrm{acc,A_x}}} + \underbrace{\int_{A_x^c} q(x,y)\alpha(x,y)|y|dy}_{=:J_{\mathrm{acc,A_x^c}}} \\ &+ \underbrace{\int_{\mathbb{R}} q(x,y)(1-\alpha(x,y))|x|dy}_{=:J_{\mathrm{rej},\mathbb{R}}}\end{aligned}$$

As shown below, Lemmas 8.3 and 8.4 imply upper bounds for $J_{\mathrm{acc,A_x}}$, $J_{\mathrm{acc,A_x^c}}$ and $J_{\mathrm{rej},\mathbb{R}}$ for sufficiently large x. Note that for arbitrary $\delta_1 > 0$ for sufficiently large x we have $Q(x, A_x^c) < \delta_1$. Let ϕ denote the density function of $N(0,1)$.

$$\begin{aligned}J_{\mathrm{acc,A_x}} &\leq \sup_{y\in A_x}|y| \int_{A_x} q(x,y)\alpha(x,y)dy \\ &\leq (1-\mathsf{a}+\frac{\varepsilon}{2})x \\ J_{\mathrm{acc,A_x^c}} &\leq \int_{A_x^c} q(x,y)|y|dy\end{aligned}$$

$$\begin{aligned}&\leq \int_{A_x^c} q(x,y)ydy - 2\sigma_x \int_{(-\infty,-\frac{\mu_x}{\sigma_x})} \phi(z)zdz \\ \int_{A_x^c} q(x,y)ydy &= \mu_x \int_{A_x^c} q(x,y)dy + \int_{A_x^c} q(x,y)(y-\mu_x)dy \\ &\leq \mu_x\delta_1 + \int_{\mathbb{R}} q(x,y)|y-\mu_x|dy = \\ &\mu_x\delta_1 + \sigma_x\sqrt{\frac{2}{\pi}}.\end{aligned}$$

Note that

$$\lim_{x\to\infty}\frac{\mu_x}{\sigma_x} = \lim_{x\to\infty} x\big[1-\frac{h}{2|\beta_+-1|}(1+\frac{\beta_+-2}{\beta_+}|x|^{-\beta_+})\big]\Big(\frac{h}{|\beta_+-1|}\frac{|x|^{2-\beta_+}}{\beta_+}\Big)^{-\frac{1}{2}} = \infty.$$

Thus for every $\delta_2 > 0$ for sufficiently large x it holds that $J_{\mathrm{acc,A_x^c}} \leq \mu_x\delta_1 +$

$\sigma_x(\sqrt{\frac{2}{\pi}}+\delta_2)$. We arrive at

$$\begin{aligned} J_{\mathrm{rej},\mathbb{R}} &= x[1-\int_{\mathbb{R}} q(x,y)\alpha(x,y)dy] \\ &\le x[1-\int_{A_x} q(x,y)\alpha(x,y)dy] \\ &= x[1-\int_{A_x} q(x,y)dy] = x\int_{A_x^c} q(x,y)dy \le x\delta_1. \end{aligned}$$

Hence

$$\int_{\mathbb{R}} P(x,y)|y|dy \le x\Big[1-\mathsf{a}+\frac{\varepsilon}{2}+(2-\mathsf{b})\delta_1+\frac{\sigma_x}{x}\big(\sqrt{\frac{2}{\pi}}+\delta_2\big)\Big].$$

Set $\gamma = 1-\mathsf{a}+\frac{\varepsilon}{2}$ and note that $\gamma\in(0,1)$. Recall that $\delta_1>0$ and $\delta_2>0$ can be chosen arbitrarily small and $\lim_{x\to\infty}\mathsf{b}=\mathsf{a}$, and $\lim_{x\to\infty}\frac{\sigma_x}{x}=0$. Hence for sufficiently large x we have $(2-\mathsf{b})\delta_1<\frac{1-\gamma}{3}$ and $\frac{\sigma_x}{x}(\sqrt{\frac{2}{\pi}}+\delta_2)<\frac{1-\gamma}{3}$. Then

$$\int_{\mathbb{R}} P(x,y)|y|dy \le \frac{2+\gamma}{3}x. \tag{8.9}$$

So far we have considered large positive x but all the above steps translate easily for x's in the other tail. For negative x, sufficiently far from the origin, we obtain (8.9) with some $\gamma_-\in(0,1)$ and x replaced by $|x|$. Recall that we chosen $V(x)=|x|+1$. Set $\lambda=\frac{3}{4}+\frac{\max(\gamma,\gamma_-)}{4}$. Then for x outside sufficiently large interval containing the origin $\int_{\mathbb{R}} P(x,y)V(y)dy\le\lambda V(x)$.

It is sufficient now to show that $\int_{\mathbb{R}} P(x,dy)V(y)$, as a function of x, is bounded on compact sets. For arbitrary x we have

$$\begin{aligned} \int_{\mathbb{R}} P(x,y)V(y)dy &= \int_{\mathbb{R}} q(x,y)\alpha(x,y)(|y|+1)dy \\ &\quad +(|x|+1)\int_{\mathbb{R}} q(x,y)(1-\alpha(x,y))dy \\ &\le \int_{\mathbb{R}} q(x,y)|y|dy+|x|+2 \\ &\le \sqrt{\frac{2}{\pi}}\sigma_x e^{-\frac{\mu_x^2}{2\sigma_x^2}}+\mu_x(1-2\Phi(-\frac{\mu_x}{\sigma_x}))+|x|+2 \end{aligned} \tag{8.10}$$

Since $\pi \in C^3, \pi(x) \neq 0, l(x) \neq 0$ for all $x \in \mathbb{R}$, μ_x and σ_x are well-defined for all $x \in \mathbb{R}$ and continuous in x. Furthermore the upper bound in (8.10) is continuous in x and hence bounded on compact sets. This finishes the proof of Theorem 7.3.

Proof of Proposition 7.7 - special case with (C2.4) satisfied

The proof follows the same steps as used in the proof of Theorem 7.3. However in case of symmetric tails we need to replace Lemma 8.3 by following Lemma 8.5 and in further computations use absolute values where necessary.

Lemma 8.5. *Suppose the assumptions of Proposition 7.7 except (C2.5) hold. Then there exists $\varepsilon > 0$ such that for all sufficiently large x*

$$\inf_{y \in A_x} \alpha^C(x, y) = 1 \quad \textit{and} \quad \inf_{y \in A_x} \alpha^{net}(x, y) = 1$$

where $A_x = [(1 - \mathsf{a} - \frac{\varepsilon}{2})x, (1 - \mathsf{a} + \frac{\varepsilon}{2})x]$.

Proof. It is sufficient to show that there exists $\eta \in (2, 4)$ such that if $h \in (\eta|\beta - 1|, 4|\beta - 1|)$ then $f(\mathsf{a}) > 0$, the rest of the proof remains as in the proof of Lemma 8.3. Consider $f'(a)$ as in (8.8). Observe that $(1 + \beta) > 0$ and the equation $(1 + \beta)a^2 + (1 - 4\beta)a + 4\beta - 4 = 0$ has two roots $a_1 = \frac{4\beta - 1 - \sqrt{17 - 8\beta}}{2(\beta+1)}$ and $a_2 = \frac{4\beta - 1 + \sqrt{17 - 8\beta}}{2(\beta+1)}$ such that $a_1 < 0$ and $1 < a_2 < 2$. Thus $f'(a) > 0$ for $a \in (1, a_2)$ and $f'(a) < 0$ when $a \in (0, 1) \cup (a_2, \infty)$. Since $f(0) = f(2) = 0$, $f(1+) = -\infty$ and $f'(2) = -\beta < 0$, there exists $\eta \in (2, 4)$ such that $f(a) > 0$ for all $a \in (\frac{\eta}{2}, 2)$ and $f(a) < 0$ for $a \in (0, 1) \cup (1, \frac{\eta}{2}) \cup (2, \infty)$. $h \in (\eta|\beta - 1|, 4|\beta - 1|)$ implies that $f(\mathsf{a}) \in (\frac{\eta}{2}, 2)$.

Now we proceed to show lack of geometric ergodicity of (net)CMALA for benchmark family of distributions with heavy tails, namely as in Theorem 7.4 and Proposition 7.7. In the proofs we apply the following Theorem 8.6.

Theorem 8.6. *[Theorem 5.1 of [RT96b]] Suppose that a Markov chain Φ is ϕ-irreducible with invariant measure π not concentrated at a single point, such*

that $P(x,\{x\})$ is a measurable function, with $\operatorname{ess\,sup}\ \mathrm{P(x,\{x\})} = 1$, where the essential supremum is taken with respect to the measure π. Then the Markov chain Φ is not geometrically ergodic.

Proof of Theorem 7.4. Note that the Markov chain of (net)CMALA is ϕ-irreducible and $q^{(net)C}(x,y)$ is continuous in x and y and so $P_{(net)C}(x,\{x\})$ is a measurable function. It is sufficient now to show that $P_{(net)C}(x,\{x\}^c)$ converges to zero when x goes to infinity and the result will follow by Theorem 8.6.

Lemma 8.7. *Suppose the assumptions of Theorem 7.4 hold. Then there exists $\varepsilon > 0$ such that for sufficiently large x, if $y \in A_x := [(1-\mathsf{a}-\frac{\varepsilon}{2})x, (1-\mathsf{a}+\frac{\varepsilon}{2})x]$ then*

$$\lim_{x\to\infty} \sup_{y\in A_x} \alpha^C(x,y) = 0, \quad \lim_{x\to\infty} \sup_{y\in A_x} \alpha^{net}(x,y) = 0,$$
$$\lim_{x\to\infty} Q_C(x,A_x) = 1, \quad \lim_{x\to\infty} Q_{net}(x,A_x) = 1.$$

Proof. Suppose $\mathsf{a} \in (0,1)$, then $0 < \lim_{x\to\infty} \frac{\mu_x^{(net)C}}{x} < 1$. Furthermore, for arbitrary $\varepsilon > 0$ such that $0, 1 \notin [1-\mathsf{a}-\frac{\varepsilon}{2}, 1-\mathsf{a}+\frac{\varepsilon}{2}]$ and sufficiently large x we have that $\inf A_x > M_+$. Thus if $h \in (0, 2|\beta_+ - 1|)$ then $f(a) < 0$ by similar computations as in the proof of Lemma 8.5. Then follow the steps of the proofs of Lemmas 8.3 and 8.4.

Suppose now that $\mathsf{a} > 1$, which implies that μ_x is negative, and $\varepsilon > 0$ such that $0, -1 \notin [1-\mathsf{a}-\frac{\varepsilon}{2}, 1-\mathsf{a}+\frac{\varepsilon}{2}]$. Then for sufficiently large x we have $\sup A_x < M_-$. Thus consider $\alpha(x,y)^C$ as in (8.4) with $\beta_y = \beta_-$. But now, there exist $l_2, l_3, l_4 > 0$ such that

$$L_1 := \sup_{y\in A_x} |x|^{\frac{\beta_- - \beta_+}{2}} \leq x^{\beta_-}$$
$$L_2 := \sup_{y\in A_x} \frac{k_-}{k_+} |1-\mathsf{d}|^{\frac{\beta_-}{2}-1} (|\beta_- - 1|\beta_-)^{\frac{1}{2}} (|\beta_+ - 1|\beta_+)^{-\frac{1}{2}} \leq l_2$$

$$L_3 := \sup_{y\in A_x} e^{|x|^{\beta_+}\cdot(1+\frac{\beta_+(\mathsf{b}-\mathsf{d})^2}{4\mathsf{a}})} \leq e^{l_3 x^{\beta_+}}$$
$$L_4 := \sup_{y\in A_x} e^{-|x|^{\beta_-}\cdot|1-\mathsf{d}|^{\beta_-}(1+\frac{\beta_-(1-(1-\mathsf{c})(1-\mathsf{d}))^2}{4\mathsf{a}_-|1-\mathsf{d}|^2})} \leq e^{-l_4 x^{\beta_-}}$$

and since $\beta_+ < \beta_-$ thus $\lim_{x\to\infty}\sup_{y\in A_x}\alpha(x,y) \leq \lim_{x\to\infty}(L_1L_2L_3L_4) = 0$. Similarly for netCMALA.

To that end let $\delta > 0$. Then by Lemma 8.7, for sufficiently large x, $\sup_{y\in A_x}\alpha(x,y) \leq \frac{\delta}{2}$ and $Q(x, A_x^c) \leq \frac{\delta}{2}$. This further implies

$$P(x,\{x\}^c) \leq \sup_{y\in A_x}\alpha(x,y)\int_{A_x} q(x,y)dy + \int_{A_x^c} q(x,y)dy \leq \delta$$

which finishes the proof.

Proof of Proposition 7.7 - special case with (C2.5) satisfied

In Lemma 8.5 it was established that $f(a) < 0$ for $a \in (0,1)\cup(1,\frac{\eta}{2})\cup(2,\infty)$. This combined with similar steps as in the proof of Lemma 8.3 gives the result.

Proof of Theorem 7.5. We concentrate only on RCMALA as the proof for netCMALA follows similar steps. This proof splits into two parts. First one considers $\nu = \delta_h$ with $h > 0$ and the second one other appropriate measures ν.

(i) RCMALA with fixed h

In each transition step of RCMALA there is positive probability that the chain will move according to RWM's transition kernel. In this paragraph we only discuss the required properties of random walk Metropolis. First note that every compact set C with $\lambda(C) > 0$ is small for P. It follows from the fact that the proposal density of RWM $q(x,y)$ is continuous in x and y and positive and so is $\pi(x)$. Thus by Lemma 1.2 of [MT+96] all compact sets (with $\lambda(C) > 0$) are small for P. Now we prove that for $V(x) = |x| + 1$ it holds that $\limsup_{x\to\pm\infty} PV \leq V(x)$. We

focus on $x \to \infty$ but similar computations apply to $x \to -\infty$. Define

$$K_{a,b} = \int_a^b q(x,y)\alpha(x,y)(y-x)dy, \quad K^{abs}_{a,b} = \int_a^b q(x,y)\alpha(x,y)(|y|-|x|)dy.$$

Then $PV = V(x) + K^{abs}_{-\infty,\infty}$ and we find upper bounds for $K^{abs}_{a,b}$. Noting symmetries in q and $||y|-|x||$ and that for sufficiently large x $\frac{\pi(y)}{\pi(x)} > 1$ when $y \in (0.5x, x)$ and $\frac{\pi(y)}{\pi(x)} < 1$ when $y \in (x, 1.5x)$ we obtain that $K^{abs}_{0.5x,1.5x} < 0$. Furthermore

$$K^{abs}_{1.5x,\infty} \leq \int_{1.5x}^{\infty} q(x,y)(y-x)dy = \frac{1}{\sqrt{2\pi}} e^{-0.125x^2} \overset{x\to\infty}{\longrightarrow} 0.$$

Since $|y|-|x| < 0$ for $y \in (0, 0.5x)$, it follows that $K^{abs}_{0,0.5x} < 0$. Observe that

$$0 \leq -K_{-\infty,0} \leq \int_{-\infty}^{0} q(x,y)(x-y)dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \overset{x\to\infty}{\longrightarrow} 0$$

$$\limsup_{x\to\infty} K^{abs}_{-\infty,0} = \limsup_{x\to\infty} [-K_{-\infty,0} - 2x \int_{-\infty}^{0} q(x,y)\alpha(x,y)dy] \leq 0.$$

Hence $\limsup_{x\to\infty} PV \leq V(x)$. Now we show that $PV(x)$ is bounded on compact sets C. Further we obtain that

$$\sup_{x\in C} PV \leq 2 + \sup_{x\in C} |x| + \sup_{x\in C} (\sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} + x(1 - 2\Phi(-x))) < \infty.$$

This completes computations for RWM and now we focus on RCMALA and use appropriate subscripts whenever we refer to CMALA and RWM kernels. Now P denotes the transition kernel of RCMALA and let kernel $\tilde{P}$ be defined as follows. Suppose $A \in \mathcal{B}(\mathbb{R})$ and $x \in \mathbb{R}$ then

$$\tilde{P}(x,A) := \mathbb{1}_{l(x)\neq 0}(x) P_{CMALA}(x,A) + \mathbb{1}_{l(x)=0}(x)\mathbb{1}_{x\in A}(x).$$

where $l(x) := \frac{\partial^2}{\partial x^2} \log \pi(x)$. Thus $P(x,A) = p\ \tilde{P}(x,A) + (1-p)\ P_{RWM}(x,A)$. Since $\lambda(D) = 0$, recall that $D = \{x : l(x) = 0\}$, thus $\pi(dx)P(x,dy) = \pi(dy)P(y,dx)$. In the main part of the proof we show that the assumptions of Theorem 8.1, which guarantee geometric ergodicity, are satisfied. First note that since $P_{RWM}(x,B) > 0$ for all $x \in \mathbb{R}$ and $B \in \mathcal{B}(\mathbb{R})$ thus also $P(x,B) > 0$

and $P^2(x, B) > 0$ which implies that the chain is π-irreducible and aperiodic. Further every compact set C with $\lambda(C) > 0$ is small for P. It follows from the fact that all compact sets (with $\lambda(C) > 0$) are small for P_{RWM}, i.e. for an arbitrary compact set C there exists a probability measure λ and $\varepsilon > 0$ such that $P_{RWM}(x, A) \geq \varepsilon\lambda(A)$ for all $A \in \mathcal{B}(\mathbb{R})$ and $x \in C$. But then also $P(x, A) \geq \hat{\varepsilon}\lambda(A)$ with $\hat{\varepsilon} := (1 - p)\varepsilon$.

Next we discuss the drift condition (8.1) with $V(x) = |x| + 1$. Consider A_x as constructed in the proof of Lemma 8.3 and suppose that x is large enough so that $A_x \cap D = \emptyset$, where D is the set of zeros of $l(x)$. Then the proof that there exists $\tilde{\lambda} \in (0, 1)$ such that $\tilde{P}V \leq \tilde{\lambda}V(x)$ for sufficiently large x follows the same steps as in case of CMALA. Further since we have already proved that $\limsup_{x\to\infty} P_{RWM}V \leq V(x)$ thus for some $\lambda \in (0, 1)$ $PV \leq \lambda V(x)$ for all large x.

Now it is left to prove that $PV(x)$ is bounded on closed intervals. We showed it already for $P_{RWM}V(x)$ so now we consider $\tilde{P}V(x)$. Let C be a sufficiently large closed interval then

$$\sup_{x\in C} \tilde{P}V \quad \leq \quad \sup C + 1 + \sup_{x\in C} V(x) + \underbrace{\sup_{x\in C\cap D^c} \int_R q(x, y)\alpha(x, y)V(y)dy}_{=:N\in[0,\infty]}$$

Further we obtain that

$$N \quad \leq \quad \sup_{x\in C} V(x) + \sup_{x\in C\cap D^c} \left(\frac{1}{\pi(x)} \int_R q(y, x)\pi(y)V(y)dy\right)$$

Since π is continuous and positive, is also bounded away from zero on compact sets. Thus it is sufficient to show that

$$\int_{\mathbb{R}} \pi(y)V(y)dy < \infty \quad \text{and} \quad \sup_{x\in C\cap D^c, y\in\mathbb{R}} q(y, x) < \infty.$$

Note that $\pi \in C$ and $V \in C$ so $\int_C \pi(y)V(y)dy < \infty$. Also $\int_{C^c} \pi(y)V(y)dy < \infty$ provided that for all $y \in C^c$ it holds that $\pi(y) \propto e^{|y|\beta_\pm}$. Now we consider

$\sup_{x\in C\cap D^c, y\in\mathbb{R}} q(y,x)$. Note that $\lim_{y\pm\infty}(\mu_y \mp \sigma_y) = \pm\infty$. Let E be an interval such that for all $y \in E^c$ either $\mu_y - \sigma_y > \sup C$ or $\mu_y + \sigma_y < \inf C$. Then $\sup_{x\in C, y\in E^c} q(y,x) \leq 1$. By assumption that $\pi \in C^3$ we obtain that $|\frac{\partial^2}{\partial y^2}(\log\pi(y))|$ is continuous and thus bounded on compact sets. Hence

$$\begin{aligned}\sup_{x\in C\cap D^c, y\in E} q(y,x)) &= \frac{1}{\sqrt{2\pi}} \sup_{x\in C\cap D^c, y\in E} (\frac{1}{\sigma_y}\, e^{-\frac{(x-\mu_y)^2}{\sigma_y^2}}) \\ &\leq \frac{1}{\sqrt{2\pi}} \sup_{y\in E} \left|\frac{\partial^2}{\partial y^2}(\log\pi(y))\right| < \infty.\end{aligned}$$

Applying Theorem 8.1, we conclude that if $h > 0$ is constant and sufficiently small and π satisfies the assumptions of Theorem 7.5 then RCMALA is geometrically ergodic.

(ii) RCMALA with randomized h.

Next we use a link between the spectral gap of Markov operator and its rate of convergence to show that the result also holds with randomized h.

Throughout the remainder of the proof all quantities refer to RCMALA with randomized h. Assumption (C2.2) implies that there exists a countable set $H := \{h_i : i \in I\}$ of points $h_i \in (0,\infty)$ such that $\nu(\{h_i\}) > 0$. Further there exists a strictly decreasing sequence $(h_n)_{n\in\mathbb{N}}$ in H with $\lim_{n\to\infty} h_n = 0$. Note that whenever we refer to RCMALA with fixed constant h we use a corresponding subscript. First observe that RCMALA transition kernel P is reversible,

$$\pi(dx)\sum_{i\in I}[P_{h_i}(x,dy)\nu(\{h_i\})] = \pi(dy)\sum_{i\in I}[P_{h_i}(y,dx)\nu(\{h_i\})]$$

and so is P^2. Similarly to the case of constant h, $P_{RWM}(x,B) > 0$ for all $x \in \mathbb{R}$ and $B \in \mathcal{B}(\mathbb{R})$ implies $P(x,B) > 0$ and $P^2(x,B) > 0$. Hence both chains are π-irreducible and aperiodic. Note that above proof for RCMALA with constant h (part (i)) combined with Definition 7.1 of geometric convergence implies that transition kernel P_h^2 gives rise for sufficiently small positive h to a geometrically

ergodic chain. Theorem 8.2 implies that $\delta^{gap}_{P^2_{h_n}} > 0$ for all sufficiently large n. Since $\delta_{P^2_{h_n}} \geq \delta^{gap}_{P^2_{h_n}}$, therefore by (8.2) we obtain that $k_{P^2_{h_n}} > 0$ for large n. Thus

$$\begin{aligned}
k_{P^2} &= \inf_{0<\pi(A)<1} k_{P^2}(A) = \inf_{0<\pi(A)<1} \frac{\int_A \pi(dx) P^2(x, A^C)}{\pi(A)\pi(A^c)} \\
&= \inf_{0<\pi(A)<1} \frac{\int_A \pi(dx) \sum_{i\in I} \sum_{j\in I} [\int_{\mathbb{R}} P_{h_i}(x, dy) P_{h_j}(y, A^C) \nu(\{h_i\})\nu(\{h_j\})]}{\pi(A)\pi(A^c)} \\
&\geq \inf_{0<\pi(A)<1} \frac{\int_A \pi(dx) \sum_{i\in I} \int_{\mathbb{R}} P_{h_i}(x, dy) P_{h_i}(y, A^C) (\nu(\{h_i\}))^2}{\pi(A)\pi(A^c)} \\
&\geq \sum_{i\in I} [\inf_{0<\pi(A)<1} k_{P^2_{h_i}}(A)(\nu(\{h_i\}))^2] \geq \sum_{i\in I} [k_{P^2_{h_i}}(\nu(\{h_i\}))^2] > 0.
\end{aligned}$$

Applying (8.2), obtain that $\delta_{P^2} > 0$. Note that by Spectral Mapping Theorem (see, e.g. [Con90], [Yos95]) $(\sigma_P)^2 = \sigma_{P^2}$. Thus $\sigma_{P^2} \in [0, 1]$ and $\delta^{gap}_{P^2} = \delta_{P^2} > 0$ which further implies that $\delta^{gap}_P > 0$. Finally by Theorem 8.2 we conclude that RCMALA under assumptions of Theorem 7.5 is geometrically ergodic.

Bibliography

[Atc06] Y.F. Atchadé. An adaptive version for the metropolis adjusted langevin algorithm with a truncated drift. *Methodology and Computing in applied Probability*, 8(2):235–254, 2006.

[BPR06] A. Beskos, O. Papaspiliopoulos, and G.O. Roberts. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077, 2006.

[BPR08] A. Beskos, O. Papaspiliopoulos, and G.O. Roberts. A factorisation of diffusion measure and finite sample path constructions. *Methodology and Computing in Applied Probability*, 10(1):85–104, 2008.

[BR05] A. Beskos and G.O. Roberts. Exact simulation of diffusions. *Annals of Applied Probability*, 15(4):2422–2444, 2005.

[BS02] Andrei N Borodin and Paavo Salminen. *Handbook of Brownian motion: facts and formulae*. Springer, 2002.

[Con90] J.B. Conway. *A course in functional analysis*. Springer, 1990.

[EM13a] Pierre Etore and Miguel Martinez. Exact simulation for solutions of one-dimensional stochastic differential equations with discontinuous drift. *ESAIM: Probability and Statistics*, 2013.

[ÉM13b] Pierre Étoré and Miguel Martinez. Exact simulation of one-dimensional stochastic differential equations involving the local time at zero of the unknown process. *Monte Carlo Methods and Applications*, 19(1):41–71, 2013.

[GC11] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.

[Has70] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97, 1970.

[HST01] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

[J+04] Galin L Jones et al. On the markov chain central limit theorem. *Probability surveys*, 1(299-320):5–1, 2004.

[Kar91] Ioannis Karatzas. *Brownian motion and stochastic calculus*, volume 113. springer, 1991.

[Ken78] John Kent. Time-reversible diffusions. *Adv. in Appl. Probab.*, 10, no. 4:819–835, 1978.

[KM09] I. Kontoyiannis and S.P. Meyn. Geometric Ergodicity and the Spectral Gap of Non-Reversible Markov Chains. *Arxiv preprint arXiv:0906.5322*, 2009.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[LS88] G.F. Lawler and A.D. Sokal. Bounds on the L2 spectrum for Markov chains and Markov processes: a generalization of Cheegers's inequality. *Transactions of the American Mathematical Society*, 309(2):557–580, 1988.

[MP10] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.

[MR11] T. Marshall and G. Roberts. An adaptive approach to langevin mcmc. *Statistics and Computing, to appear*, pages 1–17, 2011.

[MT93] S.P. Meyn and R.L Tweedie. *Markov chains and stochastic stability*. Springer London et al., 1993.

[MT+96] Kerrie L Mengersen, Richard L Tweedie, et al. Rates of convergence of the hastings and metropolis algorithms. *The Annals of Statistics*, 24(1):101–121, 1996.

[Num04] E. Nummelin. *General irreducible Markov chains and non-negative operators*. Cambridge Univ Pr, 2004.

[RC04] C.P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Verlag, 2004.

[RDF78] P.J. Rossky, J.D. Doll, and H.L. Friedman. Brownian dynamics as smart monte carlo simulation. *The Journal of Chemical Physics*, 69:4628–4633, 1978.

[RGG97] G.O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.

[RR97] G.O. Roberts and J.S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab*, 2(2):13–25, 1997.

[RR98] G.O. Roberts and J.S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268, 1998.

[RR01] G.O. Roberts and J.S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367, 2001.

[RR04] G.O. Roberts and J.S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

[RR07] G.O. Roberts and J.S. Rosenthal. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of Applied Probability*, 44(2):458, 2007.

[RR08] Gareth O Roberts and Jeffrey S Rosenthal. Variance bounding markov chains. *The Annals of Applied Probability*, pages 1201–1214, 2008.

[RR09] G.O. Roberts and J.S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

[RS02] GO Roberts and O. Stramer. Langevin diffusions and Metropolis-Hastings algorithms. *Methodology and Computing in Applied Probability*, 4(4):337–357, 2002.

[RT96a] G.O. Roberts and R.L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[RT96b] GO Roberts and RL Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.

[RW00] L Chris G Rogers and David Williams. *Diffusions, Markov Processes, and Martingales: Volume 1, Foundations*. Cambridge university press, 2000.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer, 1999.

[ST99a] O. Stramer and RL Tweedie. Langevin-Type Models I: Diffusions with Given Stationary Distributions and their Discretizations*. *Methodology and Computing in Applied Probability*, 1(3):283–306, 1999.

[ST99b] O. Stramer and RL Tweedie. Langevin-Type Models II: Self-Targeting Candidates for MCMC Algorithms*. *Methodology and Computing in Applied Probability*, 1(3):307–328, 1999.

[Tie98] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability*, 8(1):1–9, 1998.

[Ver81] A Ju Veretennikov. On strong solutions and explicit formulas for solutions of stochastic integral equations. *Sbornik: Mathematics*, 39(3):387–403, 1981.

[XSL+13] Tatiana Xifara, Chris Sherlock, Samuel Livingstone, Simon Byrne, and Mark Girolami. Langevin diffusions and the metropolis-adjusted langevin algorithm. *arXiv preprint arXiv:1309.2983*, 2013.

[Yos95] K. Yosida. *Functional analysis*. Berlin New York, Springer, 1995.