Modelling And Numerical Methods For State-Dependent Diffusions

by

Xin Yang

B.Sc., Fudan University, 2011

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in the
Department of Mathematics
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2014

APPROVAL

Name: Xin Yang

Degree: Master of Science

Title of Thesis: Modelling And Numerical Methods For State-Dependent Diffusions

Examining Committee: Dr. Steve Ruuth, Professor
Chair

Dr. Paul Tupper, Associate Professor
Senior Supervisor

Dr. David Muraki, Professor
Supervisor

Dr. Weiran Sun, Assistant Professor
Internal Examiner

Date Approved: July 29, 2014

Partial Copyright Licence

Abstract

When modelling diffusive systems with stochastic differential equations, a question about interpretations of the stochastic integral often arises. Using simulations of a random Lorentz gas model, we show that given only the diffusion coefficient, for a diffusive system without external force, the system is underdetermined. By varying one free parameter, the prediction from different interpretations can hold true. However, for a diffusive system satisfying detailed balance condition, we show that it is uniquely determined by the equilibrium distribution in addition to the diffusion coefficient. We propose an explicit method for simulating stochastic differential equations in this formulation. Our numerical scheme introduces Metropolis-Hastings step-rejections to preserve the exact equilibrium distribution and works directly with the diffusion coefficient rather than the drift coefficient. We show that the numerical scheme is weakly convergent with order $1/2$ for such systems with smooth coefficients. We perform numerical experiments demonstrating the convergence of the method for systems not covered by our theorem, including systems with discontinuous coefficients.

To my mother and father: Lin Yang and Yong Jiang
To my wife: Yue Zhao

“HAKUNA MATATA”

— The Lion King, 1994

Acknowledgments

I would like to express my sincerest appreciation to my supervisor, Professor Paul Tupper for his patience, enthusiasm, understanding and guidance at every step in my graduate study and research. At the same time, I am deeply thankful to Professor Nilima Nigam and Professor Paul Tupper for not only guiding me through the academic research but also for taking care of my life at SFU. You have set up the most important and inspiring example for me and I will spend my life to pursue this goal.
I would also like to thank Professor David Muraki and Professor Weiran Sun for reading my thesis and providing helpful comments and suggestions.
I am grateful to all the professors for giving excellent courses and to my colleagues for creating a great environment. All I learn from you will be one of the most valuable experience in my life.
Finally, thank my wife, my parents, for their uncountable love and support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Backgroud

Diffusion is a very common phenomenon and it plays an important role in many subjects. Researchers have been using various diffusion models for many processes such as particle motion, energy transportation, chemical reactions, etc. One simple diffusive model is the evolution of concentration $c(x,t)$ in one dimension under Fick's law:

$$J(x,t) = -D\frac{\partial c(x,t)}{\partial x}$$

which states that the local flux $J(x,t)$ at position x and at time t is proportional to the negative of the gradient of the concentration with coefficient D. Here, D is a constant known as the diffusion coefficient. Then, by conservation, we can obtain a partial differential equation (PDE) model for the evolution of concentration with proper initial values and boundary conditions:

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2} \tag{1.1}$$

This is a continuum model, which implies that the system contains infinitely many particles. Interestingly, it is possible to study the same process by only looking at the movement of a single particle $X(t)$ under appropriate assumptions [14]. The movement of $X(t)$ is related to the famous Brownian motion [37] and $X(t)$ is governed by a stochastic differential equation (SDE) [30] (see Appendix A.2 and A.4 for their definitions),

$$dX(t) = \sqrt{2D}dB(t) \tag{1.2}$$

As the models are describing the same physical process, it turns out that the SDE (1.2) uniquely corresponds to the PDE (1.1) which is known as the Fokker-Planck equation (FPE) for the SDE (1.2) [14] (see Appendix A.5 for the definition of the FPE). The described model is a diffusion process

with constant diffusion coefficient.

A problem arises when one starts to model a diffusion process with a spatially dependent diffusion coefficient (the diffusion coefficient is not a constant but depends on the space variable). Since the Brownian motion is not of bounded variation, for a spatially dependent diffusion coefficient $D(x)$, the Lebesgue integral of the term $\sqrt{2D(X(t))}dB(t)$ is not well defined any more. For the SDE

$$dX(t) = \sqrt{2D(X(t))}dB(t) \tag{1.3}$$

different interpretations (Ito, Stratonovich or Isothermal) of $\sqrt{2D(X(t))}dB(t)$ lead to different solutions [30]. From the PDE model, if one only modifies Fick's law to adapt to the variable diffusion coefficient $D(x)$, then the PDE (1.1) becomes the divergence form equation

$$\frac{\partial c(x,t)}{\partial t} = \frac{\partial}{\partial x} D(x) \frac{\partial c(x,t)}{\partial x}$$

However, this PDE is only equivalent to the SDE in one particular interpretation(the Isothermal interpretation, see Table 2.2). As a result, the interpretation of the stochastic term $\sqrt{2D(X(t))}dB(t)$ cannot be determined arbitrarily. In reality, a modeler often needs additional information about the process to choose the suitable interpretation. For instance, for realistic physical stochastic systems with noise, since the noise often has finite correlations and white noise is only a idealized approximation, the choice of Stratonovich integral is more appropriate [19]. However, for those applications of stochastic integrals in finance, they often require that the process does not look into the future (all decisions can only be made upon current and past information). In such situations, Itô integral seems to be a better choice [37]. In this thesis, we look at this modelling problem from a different point of view. The different behavior in the macroscopic and mesoscopic models imply that the clues about how to build the model may be obtained by looking at microscopic level models.

1.2 Thesis Organization

In Chapter 2, we rephrase the modelling problem stated in the introduction with a concrete example: a particle diffusing in a box with piecewise constant diffusion coefficients. We consider the equilibrium distribution ρ_{eq} given piecewise constant diffusion coefficient D on each side of the box. A discussion of a seeming paradox about the predictions from Statistical Mechanics and time change is presented. To resolve this paradox, we run simulations of a microscopic deterministic Hamiltonian billiard system: the random Lorentz gas [10]. The system has two free parameters: disc radius and free volume fraction. There is a one-parameter family of values of these parameters that can generate the same effective diffusion coefficient. We will show how to choose these parameters to obtain arbitrarily good approximations to diffusive motion for the particle on each side of the box. Using the degree of freedom in the choice of parameters, we show that the equilibrium density of the particle is underdetermined by the diffusion coefficient on each side of the region.

For a given diffusion coefficient, exploiting the flexibility in the parameter choice allows us to create systems in which either the statistical mechanics prediction or the time-change prediction is correct. Therefore, choosing a diffusion coefficient D is not enough to fully specify diffusive dynamics in the absence of other assumptions. In particular, our results show that fixing D is not enough to specify ρ_{eq}, the equilibrium probability density for the position of the particle. On the other hand, diffusions that satisfy the detailed balance condition with respect to some invariant measure feature prominently in many areas of physics, chemistry, and mathematical biology [45, 14, 4]. For such systems, we propose a framework for modelling state-dependent diffusion with a known equilibrium probability density. Rather than simply specifying a state-dependent diffusion coefficient, we specify diffusion coefficient $D(x)$ and an equilibrium density $\rho_{\mathsf{eq}}(x)$ which together with a detailed balance assumption (no-flux in equilibrium) completely determine the dynamics. We conclude the chapter by explaining the relation between our results and three different interpretations of state-dependent diffusion: Itô, Stratonovich and the isothermal convention [23].

In Chapter 3, we introduce a new method for the numerical simulation of diffusive dynamics which makes use of our framework: the numerical method is expressed in terms of $D(x)$ and $\rho_{\mathsf{eq}}(x)$ and makes no reference to a drift term. The method consists of Euler-Maruyama steps for a purely diffusive Itô SDE together with Metropolis rejections. The method is similar to the Metropolis- adjusted Langevin algorithm (MALA) [36],[5], except that there is no drift term in the Euler-Maruyama step, and it is the Metropolis rejections that induce any drift in the trajectories. The advantages of our method over the Euler-Maruyama method are that it samples space with the correct equilibrium density $\rho_{\mathsf{eq}}(x)$ and since our numerical method does not calculate the drift term explicitly, it is possible to handle the case when $D(x)$ and $\rho_{\mathsf{eq}}(x)$ have discontinuities where the drift would have a singularity. A proof of the convergence of the method with sufficiently smooth coefficients is provided. Then we validate the numerical methods using several examples. For these examples, the method shows that it preserves the exact equilibrium distribution and converges weakly even with discontinuous coefficients.

In Chapter 4, we establish proof of the convergence using a different technique from the point of view of Markov processes and martingale problems, which is more probabilistic in nature. As developed by Stroock and Varadhan [42], the diffusion approximations can be done through the martingale characterization in the limit. Though the key estimates in the proof in Chapter 3 cannot be avoided, this way of proof gives a more fundamental reason for the convergence. Without giving the convergence rate, the weak convergence of our numerical method is proved under conditions with less smoothness of the coefficients.

1.3 Publication and contributions of authors

The work in chapter 2 has been published in [43]: Paul Tupper developed the model and wrote the manuscript, Xin Yang implemented the the code, ran the simulations.
The work in chapter 3 has been submitted : Paul Tupper developed the numerical scheme. Xin Yang carried out the convergence proof, implemented the code and ran the simulations. Xin Yang wrote the manuscript with extensive revisions by Paul Tupper.

Chapter 2

Modelling state-dependent diffusions

2.1 A paradox of state-dependent diffusions

As shown in the introduction, specific choices need to be made when one wants to model a system with state-dependent diffusions. In order to illustrate the situation, we begin with a seeming paradox about the state-dependent diffusion.

Consider a particle diffusing in a two-dimensional box with reflecting boundary conditions. We show a portion of a simulated trajectory of such a system in Figure 2.1. Suppose that in the left half of the box the particle diffuses with coefficient D_1, and that in the right half of the box the particle diffuses with coefficient $D_2 = 2D_1$. We assume that there are not external forces acting on the particle. Our question is: does the particle spend an equal fraction of time on each side of the box in the long run?

One answer is based on the statistical mechanics: the particle will spend an equal proportion of time on each side of the box. The justification for this prediction is the principle of statistical mechanics which states that 'an isolated system in equilibrium is equally likely to be in any of its accessible states' [33, p. 54]. Since all states in the box are accessible, and there are an equal number of states on each side of the box, the particle should spend an equal proportion of its time on each side of the box.

Another answer is based on the idea of rescaling time in one side of the box: the particle will spend less time on the side of the box where the diffusion coefficient is greater. The justification for this prediction is that, in the absence of any drift, faster diffusion is equivalent to time passing more quickly. Since the geometry is reflection symmetric, for every trajectory, its reflection will also be a possible trajectory with the same probability. This means that the periods of time the particle spends

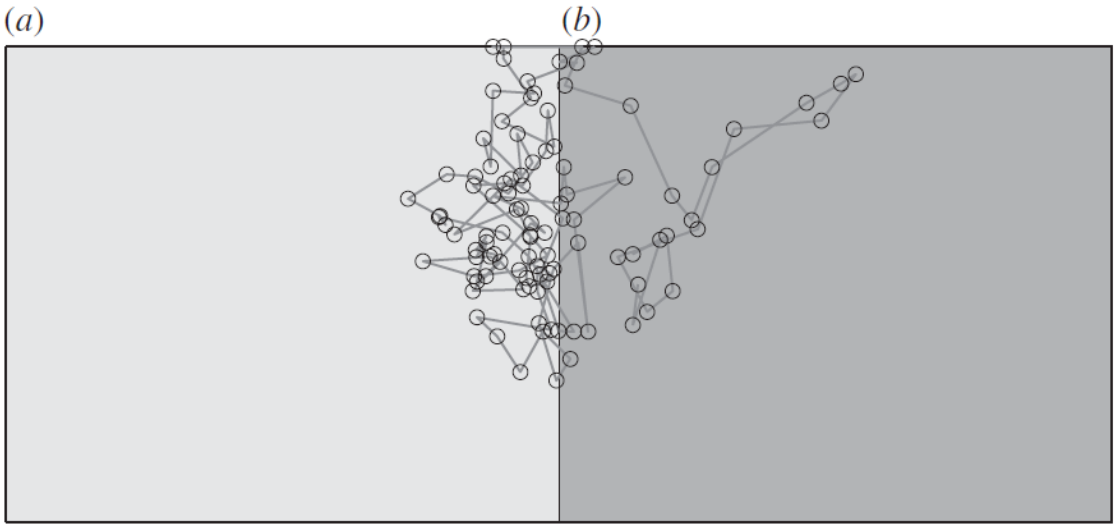

Figure 2.1: Simulation of particle diffusing within a box with reflecting boundary conditions. Points are generated by the Euler-Maruyama method $X_{n+1} = X_n + \sqrt{2hD(X_n)}N_n$, where $h = 0.5$, N_n are independent standard two-dimensional Gaussians, and $D(x)$ is the state-dependent diffusion coefficient. (a) $D(x) = 1$ and (b) $D(x) = 2$

on the right-hand side of the box will be shorter than those spends on the left hand side. Hence the total time the particle spends on the right-hand side will be less. We show how this prediction is a straightforward consequence of interpreting the particle's motion as drift-free diffusion, where we interpret the state-dependent diffusion coefficient using the Itô convention [14] A.4. We can write the equation for the particle motion as

$$dx = b(x)dB(t) \tag{2.1}$$

where $x = (x_1, x_2)$ and $B(t)$ is standard two-dimensional Brownian Motion. (Equivalently, we may write this equation as ($dx/dt = b(x)\eta(t)$) where η is two-dimensional Gaussian white noise.) We specify $b(x) = b_1$ for $x_1 < 0$ and $b(x) = b_2$ for $x_1 > 0$. Here $b_i = \sqrt{2D}$, $i = 1, 2$ where D_i is the corresponding diffusion coefficient. We enforce reflecting boundary conditions at the four walls of the box. The (Itô-)Fokker-Planck equation for $\rho(x, t)$, the probability density of the particles position at time t, is [14, p. 118]

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{1}{2}\nabla \cdot [\nabla(b^2(x)\rho(x,t))] = \nabla \cdot [\nabla(D(x)\rho(x,t))].$$

The equilibrium density $\rho_{\text{eq}}(x)$ satisfies

$$\nabla(D(x)\rho_{\text{eq}}(x)) = const.$$

Reflecting boundary conditions for the diffusion correspond to homogeneous Neumann boundary conditions $J(x,t) \cdot \vec{n} = \nabla(D(x)\rho_{\text{eq}}(x)) \cdot \vec{n} = 0$ (zero-flux) for the Fokker-Planck equation. With these

boundary conditions, the unique equilibrium density is $\rho_{\mathsf{eq}}(x) = C/D(x)$ for some constant C. Thus, since $D_2 = 2D_1$, the particle spends half as much time on the right-hand side of the box as on the left. (We discuss the relation of our question to other interpretations of (2.1) in 2.4).

Neither the statistical mechanics prediction nor the time-change prediction are definitive. The statistical mechanics prediction relies on the principle of equal probability of all accessible states, which needs to be independently justified for the mesoscopic level of description we are considering here. The time-change prediction is not definitive since Itô stochastic differential equations (SDEs) are themselves mesoscopic models whose use can only be rigorously justified by showing how they arise as the coarse-scale limit of microscopic dynamics. Since the two predictions contradict each other, at least one of them must be wrong for any given physical system. The approach by which we resolve this apparent contradiction is to study a microscopic model of the box system we have described above and then see what proportion of the time the particle spends on each side in simulations of that system. If one prediction turned out to always be true for our model system, we could use our result as a baseline for investigating under which more general situations the prediction was still true. However, we will see that even for our simple system, there are parameter choices that make the statistical mechanics prediction correct and parameter choices that make the time-change prediction correct. This demonstrates that there is no a priori reason to conclude that one prediction or another is correct, given only the diffusion coefficient $D(x)$ for the system.

2.2 A model system for state-dependent diffusion

In 2.1, we described a two-dimensional system consisting of a single particle diffusing inside a rectangular box and reflecting off the boundaries. The diffusion coefficient is twice as large on the right-hand side of the box as the left. In this section, we demonstrate how to construct a family of deterministic systems that approximates this behaviour on a coarse scale. In 2.2.1, we describe the random Lorentz gas [10], a deterministic system that when given a random initial condition yields constant-coefficient diffusion at a coarse scale. In 2.2.2, we show how to approximate the model system in 2.1 by creating two adjacent domains of the random Lorentz gas within a bounding box. In 2.2.3, we describe numerical experiments with the box system demonstrating that the fraction of time a particle spends on each side of the box cannot be determined solely from the values of the diffusion coefficient. In 2.2.4 , we explain how the result of the numerical experiments in 2.2.3 can be predicted from properties of the dynamics of the microscopic system.

2.2.1 Random Lorentz gas

Consider infinitely many discs with positions fixed in $\mathbb{R}^2$. The centers of the discs are distributed randomly with uniform density subject to the constraint that the discs do not overlap. We consider

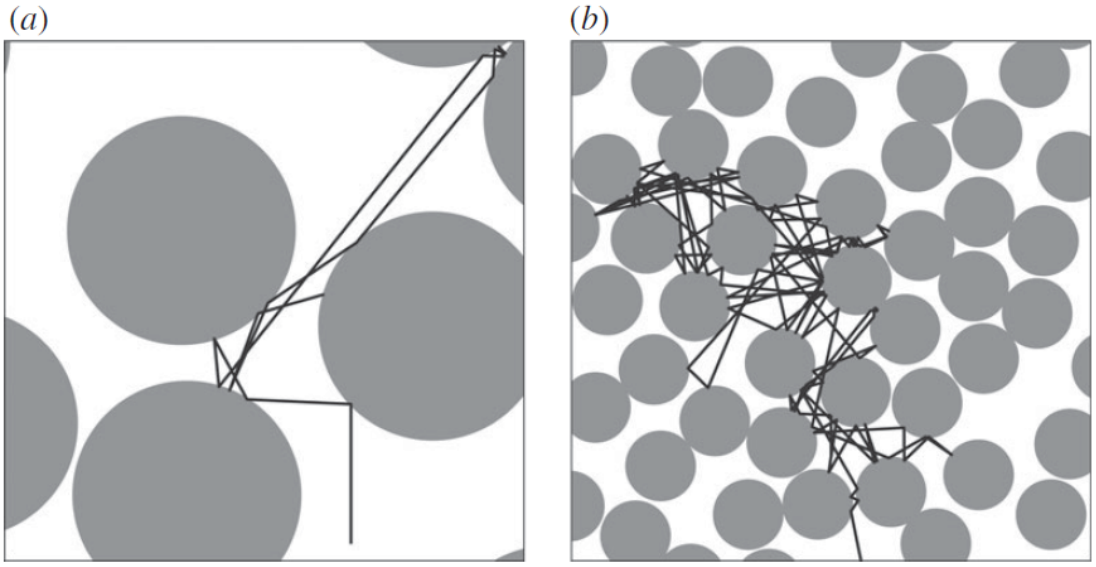

Figure 2.2: Portions of the trajectory of the random Lorentz gas with $\phi = 0.5$ at two levels of magnification, showing (a) the discs and (b) the trajectory of the moving particle.

a single point particle interacting with the discs in the following manner. Given an initial velocity and an initial position not on a disc, the particle moves with constant velocity until it meets a disc. Then it undergoes an instantaneous elastic reflection with the boundary of the disc, with the angle of incidence equalling the angle of reflection. This model is the two-dimensional random Lorentz gas [10], a mathematical formulation of a model originally due to Lorentz [26]. We will always consider the case when the particle has initial velocity of magnitude 1. Figure 2 shows trajectories of the random Lorentz gas at two different scales. The random Lorentz gas has two parameters: the radius of the discs r, the number of the discs per unit area λ. The radius r can take any positive value. The density λ has a maximum value $\lambda_{\max} = 1/(r^2\sqrt{12})$ corresponding to the close-packed hexagonal pattern of discs. We define ϕ to be the free volume fraction, the proportion of the area not occupied by discs. We have that $\phi = 1 - \pi r^2 \lambda$. The free volume can take any value in $[\phi_{\min}, 1)$ where $\phi_{\min} = 1 - \pi/\sqrt{12} \approx 0.093$, regardless of the value of r. The pair (r, ϕ) provides an alternative parametrization of the random Lorentz gas, with the advantage that ϕ is dimensionless. For $\phi = \phi_{\min}$, adjacent discs are touching and the particle remains in a small region of the plane for its entire trajectory. For $\phi \in (\phi_{\min}, 1)$, the probability that two discs are touching anywhere in the plane is zero [10, p. 325] and motion of the particle is conjectured to be diffusive [10]. Specifically, if we start the particle at initial position $x(0)$ not coincident with a disc and give it initial speed 1 and uniformly distributed direction in $[0, 2\pi)$, then $x(t)$, the position of the particle at time t is approximately distributed as a Gaussian random vector with mean 0 and variance matrix $2DtI$. Here I is the 2×2

identity matrix and D denotes the diffusion coefficient. This conjecture is supported by analytical calculations [11][10] and numerical simulations [6][9].
A stronger and more formal statement of the conjecture is stated in the language of weak convergence [2]. Specifically, let $x(t)$ denote the position of the moving particle at time t . Fix $x(0)$ to be some point not coincident with a disc and let the initial velocity be chosen as above. It is conjectured that

$$\frac{x(nt)}{\sqrt{n}} \Rightarrow \sqrt{2D_{r,\phi}}B(t)$$

as $n \to \infty$, where $B(t)$ is standard two-dimensional Brownian motion , $D_{r,\phi}$ is the diffusion coefficient which depends on r and ϕ, and $\Rightarrow$ denotes weak convergence in the space of continuous functions [2]. Such a result holds for the certain periodic Lorentz gasses [7] [35], but remains open for the random Lorentz gas. (We chose for our study not to use the standard periodic Lorentz gas with discs centred on a hexagonal lattice since for large enough ϕ the particle undergoes superdiffusive motion in this case.) A scaling argument shows that, for fixed ϕ, $D_{r,f}$ is proportional to r . To see this, letting $|\cdot|$ denote the Euclidean norm and letting $\langle\cdot\rangle$ denote the average, observe that the coefficient D can be obtained by the Einstein-Smoluchowsky equation,

$$D = \lim_{t\to\infty} \frac{\langle|x(t)|^2\rangle}{2td},$$

where $d = 2$ is the dimension in this case. If we rescale space by a factor R, we increase both the size of the discs and the distance the particle travels by a factor of R, without changing ϕ. So $\langle|x(t)|^2\rangle$ increases by a factor of R^2 . To maintain the speed of the particle as 1, we also have to rescale time, increasing t by a factor R. The net effect on the ratio $\langle|x(t)|^2\rangle/4t$ is to increase it by a factor R [38]. So

$$D_{r,\phi} = rf(\phi) \tag{2.2}$$

for some function f of ϕ. Figure 2.3 shows the relation between $f(\phi)$ and ϕ that was computed using the techniques described in [9]. It appears that the function $f(\phi)$ is continuous on its domain, it is monotonically increasing, $f(\phi) \to 0$ as $\phi \to \phi_{\min}$ and $f(\phi)$ goes to infinity as $\phi \to 1$. Indeed, calculations from kinetic theory show that $f(\phi) \sim 3\pi/[16(1-\phi)]$ in the $\phi \to 1$ limit [46] [6]. Given any fixed diffusion coefficient $D > 0$ there is a one-parameter family of choices of r, ϕ such that $D = D_{r,\phi}$: for any $\phi \in (\phi_{\min}, 1)$, just choose $r = D/f(\phi)$.

2.2.2 Box with two domains

In order to investigate the main question about the different predictions, we take a rectangular box and divide it into two equal regions. Each region is filled with randomly placed discs, with different r and ϕ on each side. As in the Lorentz gas, the centers of the discs are placed uniformly at random

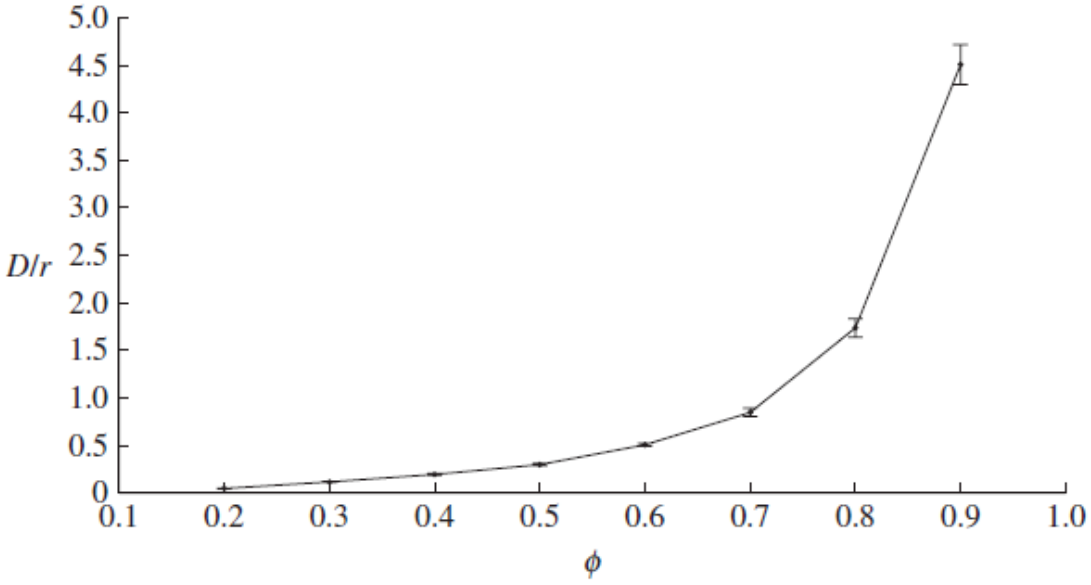

Figure 2.3: The ratio of diffusion coefficient versus disc radius D/r as a function of free volume fraction ϕ for the random Lorentz gas.

with the condition that they not overlap with each other. Discs can intersect with the walls of the box but not the dividing line between the two sides of the box. Figures 2.4 and 2.5 show two examples of such configurations of discs. The dynamics of the point particle are the same as in the infinite random Lorentz gas, with the added condition that the particle reflects off the walls of the box. Suppose we are given diffusion coefficients $D_1, D_2 > 0$. We can choose r_1, ϕ_1 and r_2, ϕ_2 such that $D_1 = D_{r_1,\phi_1}$ and $D_2 = D_{r_2,\phi_2}$. For small enough r_1, r_2 , the dynamics of the particle in this system will be well-approximated by a particle that diffuses with coefficient D_1 on the left-hand side of the box and diffuses with coefficient D_2 on the right-hand side of the box. With appropriate choices of the parameters, we can investigate the question of the proportion of time the particle spends on each side of the box.

2.2.3 Numerical experiments

We start the particle off at some position in the box not on a disc with velocity of magnitude 1 and randomly chosen direction. The motion of the particle is simulated with an event-driven simulation, computing a trajectory that is accurate up to the errors of floating point arithmetic [9]. Periodically (with a fixed time interval Δt) the position of the particle is recorded. At the end of a long trajectory, we compute the number of times the particle is on the right-hand side of the box divided by the number of times the particle is on the left-hand side of the box. We consider two choices of

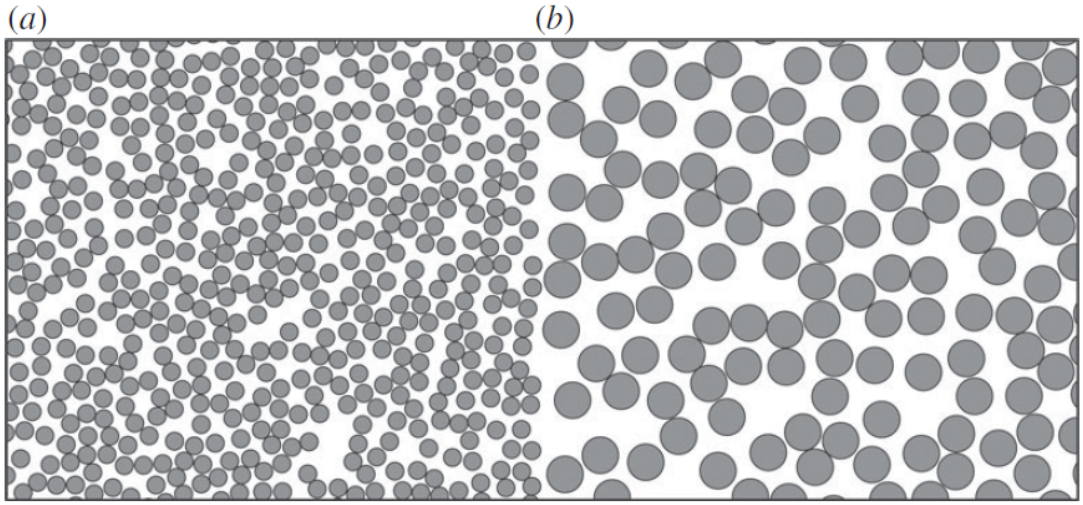

Figure 2.4: Set-up 1. Free volume fraction is the same on each side $\phi_1 = \phi_2$. Scatterer radius on right (b) is twice that on left (a): $2r_1 = r_2$, leading to $2D_1 = D_2$

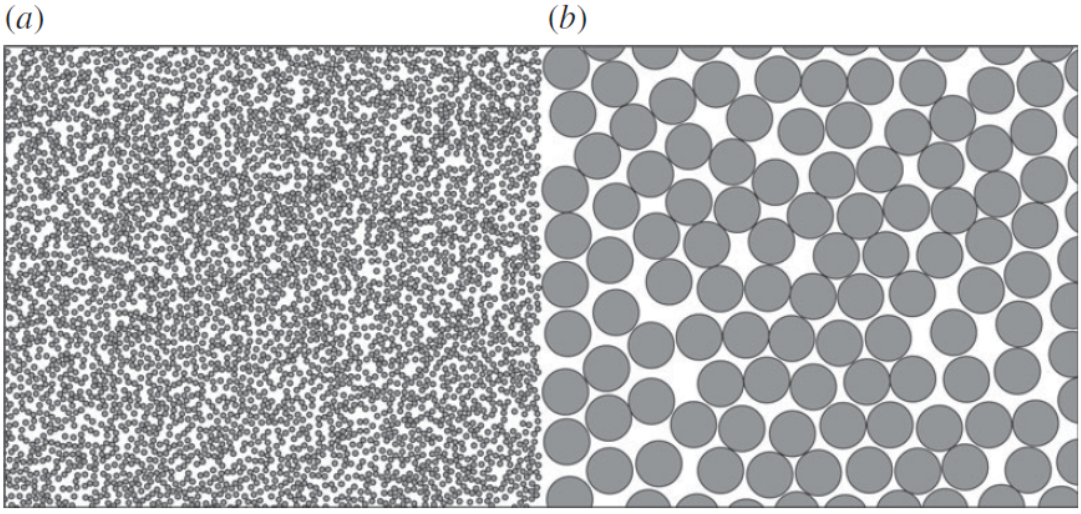

Figure 2.5: Set-up 2. Free volume fraction is twice as big on left (a): $\phi_1 = 2\phi_2$. Scatterer radii are chosen so that $2D_1 = D_2$

set-up	r_1	ϕ_1	D_1	r_2	ϕ_2	D_2	time on the right/time on the left
1	0.3	0.5	0.09	0.6	0.5	0.18	0.99
2	0.09	0.60	0.47	0.75	0.30	0.093	0.50

Table 2.1: The parameters used in each side of the box in the two set-ups, and the ratio between amount of time spent by the particle on the right-hand side of the box and the left-hand side of the box in each set up

parameters on each side of the box, or set-ups, each of which leads to the effective diffusion coefficients satisfying $D_2 = 2D_1$. We have contrived the set-ups so that in set-up 1 in Figure 2.4, the statistical mechanics prediction is correct, and that in set-up 2 in Figure 2.4, the Time-change prediction is correct. The parameters for each set-up are summarized in Table 2.1. In the first set-up $\phi_1 = \phi_2 = 0.5$ and $2r_1 = r_2 = 0.6$. We show the position of the discs in the box in figure 2.4. Over a trajectory of length 5×10^5 time units the ratio between the occupation times is approximately 1, as we show in table 2.1. This result agrees with the statistical mechanics prediction.
In the second set-up $2\phi_2 = \phi_1 = 0.60$ and $8.3r_1 \approx r_2 = 0.75$. We show the position of the discs in the box in Figure 2.5. Over a trajectory of length 7×10^6 time units the ratio between the occupation times is approximately $\frac{1}{2}$, as we show in table 2.1. This result agrees with the time-change prediction. Thus, by fixing the parameters appropriately, both the statistical mechanics prediction and the time-change prediction can be seen to be correct for the given mesoscopic behavior.

2.2.4 Analysis and discussion

We first explain that, given the properties of random Lorentz gas, the results of the simulation in the previous subsection are predictable. We first note that the system is ergodic. The ergodicity of periodic Lorentz gas is shown by [40], since it is equivalent to dispersing billiards on the torus. Our system is not dispersing because the walls of the box are not convex, but [40, Section 9] shows how ergodicity still holds in this case by unfolding the box to obtain a periodic domain. Ergodicity implies that the duration of time that the particle spends in a region of phase space is proportional to the volume of the region. For our system, this implies that the amount of time spent by the particle on each side is proportional to the free volume fraction ϕ on each side. For fixed ϕ, the parameters r and D are irrelevant to the proportion of time the particle spends on each side of the box. In set-up 1 Figure 2.4 above, $\phi_1 = \phi_2$ so the ratio of times spent on each side of the box is equal and the statistical mechanics prediction is correct. In set-up 2, $\phi_1 = 2\phi_2$, and so the particle spends twice as much time on the left-hand side of the box, and the time-change prediction is correct. Despite appearances, the principles of statistical mechanics are not violated in set-up 2. There are two ways to reconcile the apparent contradiction. Firstly, one can say that at the microscopic scale statistical mechanics is not violated because there are not an equal number of states on each side. The number of states is proportional to the free volume fraction on both sides, and so the system spends

more time where the free volume fraction is higher. The other way of reconciling the disagreement is at the mesoscale. Suppose we divide the box into many rectangular cells of equal area, each much smaller than the whole box, but much larger than the size of the discs. Each cell corresponds to a mesoscopic state which the particle may be in. Since the system is in the microcanonical ensemble (no exchange of energy with the outside) the probability of the system in equilibrium being in a particular mesoscopic state is determined by the entropy of that state, where the entropy of a mesoscopic state is proportional to the logarithm of the amount of microscopic states it contains. The greater free volume fraction on the left in set-up 2 implies that mesoscopic states have greater entropy there, and so the system will spend more time on the left-hand side of the box. On the other hand, in set-up 1, the time-change prediction proves to be wrong. This means that the motion of the particle in the box is not well-described by the drift-free Itô SDE (2.1). Since, by the properties of the uniform random Lorentz gas, (2.1) is a good model for the dynamics of the particle within each of the two regions of constant disc radius r , it must be that the equation is no longer a good model at the boundary of the two regions. We see that there are choices of ϕ_1 ,r_1 and ϕ_2 ,r_2 such that either the Time-change prediction or the statistical mechanics prediction are correct, while still $D_2 = 2D_1$. Indeed, for any D_1 and D_2 , parameters can be chosen to induce arbitrary ratios between the times spent on the left-hand side and the right-hand side. Although for generic values of the parameters in our Lorentz gas model neither prediction will be valid, we point out that the statistical mechanics prediction holds for a natural set of parameter settings, whereas the same is not true of the time-change prediction. For the time-change prediction to be correct, it is necessary for the ratio between ϕ_1 and ϕ_2 to match the ratio between D_2 and D_1 . We see no natural way that the parameters in our model may be set for this matching to occur. On the other hand, for the statistical mechanics prediction to be correct it is necessary that ϕ_1 and ϕ_2 be equal. There is at least one case where this condition approximately holds for a naturally occurring system. Consider a situation in which both $\phi_1, \phi_2 \approx 1$. This requires no fine tuning, only that the discs take up a small fraction of the total volume. Choosing r_1 and r_2 to be unequal leads to different diffusion coefficients on each side, but the particle still spends approximately equal proportions of time in each region. Likewise, in any physical system where a particle diffuses by interacting with small, sparsely placed scatterers, we expect the statistical mechanics prediction to be correct.

Our justification for studying a particular microscopic system is threefold. Firstly, there is a solid analytical understanding of the random Lorentz gas on which we can base our simulations. Secondly, our purpose is not to conclude that a particular style of mesoscopic modelling is always the correct one, but to show that postulating a state-dependent diffusion rate is not enough to fully specify mesoscopic behavior. For this objective, it is enough to show that multiple mesoscopic behaviors are possible for a single class of simple models, as we do in this chapter. Finally, we have resorted to a purely deterministic microscopic model, rather than a stochastic microscopic model (such as

Langevin dynamics or a random walk) to avoid concerns that the method used to introduce randomness at the microscopic level somehow biases the results at the mesoscopic level. This last point distinguishes our work from similar discussions of [45] and [31]. These authors show that spatially inhomogeneous random walks, with natural choices of parameters, can lead to either prediction holding true at the mesoscopic level. Similarly, our conclusions closely parallel those of [21], in which the same question is answered using a random-walk model on a one-dimensional lattice. There, following the earlier work of [32], they show that choosing how the random walk behaves at the interface of two regions of differing diffusion coefficient leads to different equilibrium probability densities for the system. In [21], the authors explain how to determine the correct interface behavior of the random walk model using experimentally measurable quantities. Our approach differs in that, in our model, the interface behavior is determined indirectly through the microscopic dynamics that we describe.

2.3 A proposal for modelling with state-dependent diffusion

The numerical simulation in 2.2.3 demonstrates that the equilibrium density $\rho_{\mathsf{eq}}(x)$ is not determined solely by the local diffusion rate $D(x)$, even in situations with no external forces acting on the particle. This raises the practical issue of how to model systems with state-dependent diffusion. Ideally, mesoscopic diffusive models would be derived from microscopic models via an asymptotic technique, such as the van Kampen system-size expansion[45, ch. XI.3]. But in many circumstances, deriving a realistic microscopic model may be impractical. Instead, in this section, we describe a more phenomenological approach. We assume that the modeller posits an isotropic state-dependent diffusion rate $D(x)$ and an equilibrium density $\rho_{\mathsf{eq}}(x)$. We then derive a drift coefficient $a(x)$ for an Itô SDE with diffusion coefficient $D(x)$ that gives the desired $\rho_{\mathsf{eq}}(x)$. In 2.3.2, we show how the Lorentz gas system of 2.2.3 can be modelled at a mesoscopic level in this way.

2.3.1 A modelling framework for state-dependent diffusion

We model the diffusion in k spatial dimensions by the Itô SDE

$$dX(t) = a(X(t))dt + \sqrt{2D(X(t))}dB(t), \tag{2.3}$$

where we will determine $a(x)$ in terms of $\rho_{\mathsf{eq}}(x)$ and $D(x)$. Here $B(t)$ is standard k-dimensional Brownian motion. (In alternate notation, we write this equation as $dX(t)/dt = a(X(t)) + \sqrt{2D(X(t))}\eta(t)$ where η is k-dimensional Gaussian white noise.) The Fokker-Planck equation for this system is

$$\frac{\partial}{\partial t}\rho(x,t) = -\nabla \cdot [a(x)\rho(x,t)] + \Delta[D(x)\rho(x,t)] = -\nabla \cdot J(x) \tag{2.4}$$

where $J(x)$ is the probability flux. Since the problem is underdetermined as stated, we need to impose extra constraints. The constraint that we consider is the detailed balance condition which

implies that the probability flux vanishes in equilibrium. This is one conclusion of the detailed balance condition from a macroscopic (PDE) point of view. A formal description of the detailed balance condition is in Section 3.3. One typical type of systems is a closed, isolated physical system where the solution satisfying detailed balance condition is known to be the thermal equilibrium distribution (see [45, ch V.6, ch. XI.4] for precise criteria on systems under which this condition holds). We choose $a(x)$ so that $J(x)$ is zero in equilibrium:

$$J(x) = a(x)\rho_{\mathsf{eq}}(x) + \nabla[D(x)\rho_{\mathsf{eq}}(x)] = 0.$$

Solving for $a(x)$ gives

$$\begin{aligned} a(x) &= \frac{1}{\rho_{\mathsf{eq}}(x)}\nabla(D(x)\rho_{\mathsf{eq}}(x)) \\ &= \nabla D(x) + D(x)\nabla \ln \rho_{\mathsf{eq}}(x). \end{aligned} \tag{2.5}$$

Thus given an equilibrium density $\rho_{\mathsf{eq}}(x)$ and diffusion coefficient $D(x)$ the appropriate Itô SDE is

$$dX(t) = (\nabla D(X(t)) + D(X(t))\nabla \ln \rho_{\mathsf{eq}}(X(t)))\, dt + \sqrt{2D(X(t))}dB(t). \tag{2.6}$$

The Fokker-Planck equation of (2.6) is

$$\frac{\partial}{\partial t}\rho(x,t) = \nabla \cdot \left[-\nabla(D(x)\rho_{\mathsf{eq}}(x)) \left(\frac{\rho(x,t)}{\rho_{\mathsf{eq}}(x)}\right) + \nabla(D(x)\rho(x,t))\right] \tag{2.7}$$

which is the special isotropic case of [45, XI.4.14]. Therefore we have a formulation in terms of D and ρ_{eq} only. The advantages of this change of perspective are two-fold: (i) in many circumstances it is more natural to model the system in terms of ρ_{eq} and D, such as when ρ_{eq} is available from experimental data but a is not [39], (ii) there are situations in which D and ρ_{eq} are well-defined but a is singular, such as when D or ρ_{eq} has a jump discontinuity. In this case, defining algorithms in terms of D and ρ_{eq} allows us to avoid working with a singular drift a. Note that if $\rho_{\mathsf{eq}}(x)$ is constant with respect to x, which corresponds to the Statistical Mechanics Prediction being true, then $a(x) = \nabla D(x)$ and (2.6) reduces to

$$dX(t) = \nabla D(X(t))dt + \sqrt{2D(X(t))}dB(t).$$

On the other hand, to obtain a drift-free Itô SDE in this framework requires $D(x)\rho_{\mathsf{eq}}(x)$ to be constant in x, which means that $\rho_{\mathsf{eq}}(x)$ is determined completely by $D(x)$.

2.3.2 Connection to the Lorentz gas model

We explain the connection between the framework of 2.3 and the microscopic Lorentz gas model of 2.2. Given an instance of our random Lorentz gas model with two domains, what are the corresponding functions $D(x)$, $\rho_{\mathsf{eq}}(x)$ in (2.7), the mesoscopic equation for ρ? Suppose on the left-hand

side of the box the Lorentz gas model has parameters r_1 and ϕ_1 and on the right it has parameters r_2 and ϕ_2 . The diffusion coefficients on each side, D_1 and D_2 are determined by the relation (2.2). To determine the equilibrium probability density of the particle on each side, let $2A$ be the total area of the box, so that each side has area A. We know that $\rho_{eq,i}$, the probability density on side i, is proportional to ϕ_i , and thus $\rho_{eq,1}/\rho_{eq,2} = \phi_1/\phi_2$. We also know that since the total probability must be 1, $\rho_{eq,1}A + \rho_{eq,2}A = 1$. Solving for $\rho_{eq,i}$ gives

$$\rho_{eq,i} = \frac{\phi_i}{A(\phi_1 + \phi_2)},$$

for $i = 1, 2$. We define $D(x)$ and $\rho_{\mathsf{eq}}(x)$ for $x = (x_1, x_2)$ in the box by

$$D(x) = \begin{cases} D_1, & \text{for } x_1 < 0, \\ D_2, & \text{for } x_1 > 0, \end{cases} \qquad \rho_{\mathsf{eq}}(x) = \begin{cases} \rho_{eq,1}, & \text{for } x_1 < 0, \\ \rho_{eq,2}, & \text{for } x_1 > 0. \end{cases}$$

These functions determine $a(x)$, the drift coefficient in (2.3), via (2.5). The drift $a(x)$ is zero everywhere except along the line $x_1 = 0$ where it is not defined. We determine the appropriate boundary conditions for ρ along the boundary line $x_1 = 0$. In order for the right-hand side of (2.7) to be well defined, we require $\rho(x)/\rho_{\mathsf{eq}}(x)$ to be continuous. So for any x along the boundary line we must have

$$\frac{\rho(x^-)}{\rho_{\mathsf{eq}}(x^-)} = \frac{\rho(x^+)}{\rho_{\mathsf{eq}}(x^+)}$$

where x^- denotes taking the limit from the left, and x^+ denotes taking the limit from the right. For our particular choice of ρ_{eq} , this gives

$$\frac{\rho(x^-)}{\rho_{eq,1}} = \frac{\rho(x^+)}{\rho_{eq,2}}$$

The second boundary condition comes from assuming continuous flux across the boundary line:

$$D(x^-)\rho_{\mathsf{eq}}(x^-)\frac{\partial}{\partial x_1}\frac{\rho(x^-)}{\rho_{\mathsf{eq}}(x-)} = D(x^+)\rho_{\mathsf{eq}}(x^+)\frac{\partial}{\partial x_1}\frac{\rho(x^+)}{\rho_{\mathsf{eq}}(x+)}$$

For our particular choices of D and ρ_{eq} , since ρ_{eq} is constant away from the line $x_1 = 0$, this gives the boundary conditions

$$D_1\frac{\partial}{\partial x_1}\rho(x^-) = D_2\frac{\partial}{\partial x_1}\rho(x^+)$$

A possible direction for further investigation is to consider Lorentz gas models where disc radius r and free volume fraction ϕ vary smoothly with x , and to determine what $D(x)$ and $\rho_{\mathsf{eq}}(x)$, and hence $a(x)$.

2.4 Discussion: Itô, Stratonovich, or Isothermal

We conclude by discussing our results in the context of the apparent ambiguity between Itô, Stratonovich, and Isothermal interpretations of stochastic integrals [24][47]. For the purposes of discussion, we consider a particle moving in one spatial dimension whose position at time t is $X(t)$. We assume that $X(t)$ is a Markov stochastic process with continuous sample paths. We model the motion of the particle with the stochastic differential equation (SDE)

$$dX(t) = a(X(t))dt + b(X(t))dB \tag{2.8}$$

where $B(t)$ is standard Brownian motion. We call $a(x)$ the drift and $b(x)$ the diffusion of the SDE.

As is well known [45, 14], unless we specify a particular interpretation, the SDE (2.8) does not unambiguously define the stochastic process $X(t)$. To see this, we integrate (2.8) over $[0, T]$ to get

$$X(t) - X(0) = \int_0^T a(X(t))dt + \int_0^T b(X(t))dB(t).$$

The first term on the right has a unique interpretation as a Riemann integral, but the second term cannot be simply viewed as a Riemann-Stieltjes integral, since $B(t)$ is not of bounded variation. If we compute the second term as the limit of Riemann sums, the answer depends on where in each subinterval the argument $b(X(t))$ is evaluated. For example, choosing $h = T/N$, $t_n = hn$ and letting $B_n = B(t_n)$ and $X_n = X(t_n)$, suppose we take the integral with respect to B to be

$$\int_0^T b(X(t))dB(t) = \lim_{h\to 0} \sum_{n=0}^{N-1} b(X_n^*)(B_{n+1} - B_n),$$

where

$$X_n^* = (1-\alpha)X_n + \alpha X_{n+1}. \tag{2.9}$$

Famously, unless $b(x)$ is a constant, the limit depends on the choice of α [47]. If we choose $\alpha = 0$, we obtain the Itô interpretation of the integral, which yields a stochastic process $X(t)$ with Fokker-Planck equation

$$\frac{\partial}{\partial t}\rho(x,t) = -\frac{\partial}{\partial x}\left[a(x)\rho(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x)^2\rho(x,t)\right].$$

If we choose $\alpha = 1/2$, we obtain the Stratonovich interpretation of the integral, which yields a process $X(t)$ with Fokker-Planck equation

$$\frac{\partial}{\partial t}\rho(x,t) = -\frac{\partial}{\partial x}\left[a(x)\rho(x,t) + \frac{1}{2}b(x)b'(x)\rho(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x)^2\rho(x,t)\right].$$

Note that the Fokker-Planck equation shows that the Stratonovich interpretation of the SDE with drift $a(x)$ and diffusion $b(x)$ yields the same stochastic process as the Itô interpretation of the SDE

α	Interpretations of Stochastic Integral	Fokker-Planck Equation for $a \equiv 0$	Special Properties when $a \equiv 0$
0	Itô	$\frac{\partial}{\partial t}\rho = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b^2\rho\right]$	$\langle X(t) - X(0)\rangle = 0$ for all t.
1/2	Stratonovich	$\frac{\partial}{\partial t}\rho = \frac{1}{2}\frac{\partial}{\partial x}\left\{b\frac{\partial}{\partial x}\left[b\rho\right]\right\}$	Standard chain rule applies
1	Isothermal	$\frac{\partial}{\partial t}\rho = \frac{1}{2}\frac{\partial}{\partial x}\left\{b^2\frac{\partial}{\partial x}\rho\right\}$	$\rho_{\text{eq}} = \text{const.}$

Table 2.2: Summary of some properties of the Itô, Stratonovich and isothermal interpretations

with drift $a(x) + b(x)b'(x)/2$ and diffusion $b(x)$ [14, p. 99]. Finally, if we choose $\alpha = 1$ we obtain the anti-Itô or Isothermal interpretation [24, 47], which yields a process with Fokker-Planck equation

$$\frac{\partial}{\partial t}\rho(x,t) \quad = \quad -\frac{\partial}{\partial x}\left[a(x)\rho(x,t) + b(x)b'(x)\rho(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x)^2\rho(x,t)\right].$$

In this case, the Isothermal interpretation of the SDE with drift $a(x)$ and diffusion $b(x)$ gives the same stochastic process as the Itô interpretation of the SDE with drift $a(x) + b(x)b'(x)$ and diffusion $b(x)$ [24].

As we can see in the various Fokker-Planck equations above, if we fix $a(x)$ and $b(x)$, varying the parameter α gives different stochastic processes for the motion of the particle. This fact may make it seem like there should be a physically correct choice of the parameter α. We argue that this is misleading. For any fixed α the range of stochastic processes that can be captured by an appropriate choice of $a(x)$ and $b(x)$ is the same. For example, suppose we fix a choice of $a(x)$ and $b(x)$ and choose to interpret (2.8) with a given $\alpha \in [0, 1]$. The process defined is identical to what we would obtain with the Itô interpretation ($\alpha = 0$) of the SDE with drift $a(x) + \alpha b'(x)b(x)$ and diffusion $b(x)$.

Though the families of stochastic processes described using each convention are the same, it may still be the case that some choice of α is more natural or convenient for some purposes than others. Frequently, the rationale is based on the idea that if the drift is zero, then $X(t)$ should have certain properties. For example, if one wants $\langle X(t) - X(0)\rangle = 0$ for all t when $a(x) \equiv 0$, regardless of $b(x)$, then the Itô convention with $\alpha = 0$ in (2.9) guarantees this. If one wants that when $a(x) \equiv 0$ the equilibrium density is constant, then the Isothermal convention with $\alpha = 1$ in (2.9) guarantees this. We summarize the properties of the various interpretations of the SDE (2.8) when $a(x) \equiv 0$ in the following table. The question we posed in Section 2.1 may be rephrased as follows: in the absence of external forces, and given a diffusion $b(x) = \sqrt{2D(x)}$, what is the correct choice of parameter α and drift $a(x)$ to model the motion of the particle? A natural way to approach the problem is to interpret the absence of external forces as meaning that $a(x) \equiv 0$. Then the problem boils down to the choice of α: the Statistical Mechanics Prediction follows from taking $\alpha = 1$, and the Time-Change

Prediction follows from taking $\alpha = 0$. The results in Section 2.2.4 showed that neither answer is justified universally.

In Section 2.3 we recommended a different approach. We fix α and then choose $a(x)$ to generate the desired equilibrium distribution. As we have explained here, the choice of α is not crucial once we allow a non-zero $a(x)$. Accordingly, we have chosen $\alpha = 0$, corresponding to Itô calculus. This is the main choice in the mathematics literature, and numerical methods such as the Euler-Maruyama method take a particularly simple form with it. Once we have made this choice of α, we are free to choose $a(x)$ appropriately. In Section 2.3 we chose $a(x)$ to ensure a given equilibrium density.

Beyond the particular needs of the present work, we believe the framework we describe in Section 2.3 provides a natural and flexible way to model diffusive systems. In situations where a researcher is confident for physical reasons that the equilibrium probability is constant, then our framework takes a simple form. One area in which the proposed framework in Section 2.3 could be used is cellular biology. Although earlier models of chemistry in the cytoplasm of cells assumed that chemical species were well-mixed and thus ignored diffusion, more recent models have taken the geometry of the cell and the diffusion coefficient of various molecules into account [44]. Effective diffusion coefficients of a molecule in a cell differ from that in water owing to the crowding effect of other molecules. One approach is to model the motion of a molecule as diffusion with constant coefficient, but then have the effective diffusion be modified by interaction with other particles which are also included in the model [34]. Another is to not include the crowding particles in the model, but to model their effect with a modified diffusion coefficient [17]. Given the inhomogeneity of the cytoplasm, we can expect that this effective diffusion coefficient of the molecule (and its equilibrium probability density) will vary with location within the cell. Our framework and numerical method are developed with this latter situation in mind.

Chapter 3

A Metropolized integator for SDEs

We shall propose a numerical method for SDEs which utilizes the formulation (2.6) in Chapter 2. Before we get to the new numerical method, let's have a brief introduction to numerical methods for SDEs and Metropolis-Hastings Algorithm. A detailed introduction could be found in [18] and [8].

3.1 Numerical Methods for solving SDEs

3.1.1 The Euler-Maruyama Method

Like solving ODEs and PDEs, in most cases, analytic solutions cannot be obtained for SDEs. Hence people seek numerical methods for computer simulations. The simplest numerical method for SDEs is the Euler-Maruyama method, which can be treated as an extension of the forward Euler method for time dependent ODEs. Consider an SDE in one dimension

$$dX = a(X,t)dt + b(X,t)dB(t)$$

or in the integral form

$$X(t) = X(0) + \int_0^t a(X(s),s)ds + \int_0^t b(X(s),s)dB(s) \tag{3.1}$$

where $B(t)$ is the one dimensional Brownian motion with initial condition $X(0) = x_0$. If we approximate the integrands by $a(X(0),0)$ and $b(X(0),0)$, then we obtain the Euler-Maruyama method. Let h be the time step length and X_n be the numerical solution at $t = nh$. The Euler-Maruyama method takes an explicit iteration

$$X_{n+1} = X_n + a(X_n,t_n)h + b(X_n,t_n)\left(B\left((n+1)h\right) - B\left(nh\right)\right)$$

The term $\left(B\left((n+1)h\right) - B\left(nh\right)\right)$ could be calculated by a Gaussian random variable with mean 0 and variance h. In matlab, this is simply `dB=sqrt(h)*randn`. Higher accuracy methods can be

obtained by including higher order terms in the stochastic expansion of the integrands in (3.1). We discuss the concept of accuracy and convergence in 3.1.2.

3.1.2 Modes of convergence: Strong convergence and weak convergence of numerical methods

The way to measure error in numerical methods for SDEs is more complicated than that for ODEs. In the SDE case, we need to measure the difference between random variables or stochastic processes. Strong convergence and weak convergence are two types of convergence that are frequently used for numerical methods for SDEs. If X_n is the numerical solution from some numerical method approximating the exact solution $X(nh)$, then

$$e_h^{strong} := \mathbb{E}\,|X_n - X(nh)|$$

is the strong error. If we have

$$e_h^{strong} \leq Ch^p$$

uniformly for all $nh < T$, for some constants C, $p > 0$ when h is sufficiently small, then we call the numerical method is strongly convergent with order of accuracy p. Let

$$e_h^{weak} := |\mathbb{E}(f(X_n) - f(X(nh)))|$$

be the weak error where f is from some class of test functions which are often chosen from polynomials. If we have

$$e_h^{weak} \leq C_f h^p$$

uniformly for all $nh < T$, for some constant $p > 0$ and C_f depending on the test function f only when h is sufficiently small, then we call the numerical method is weakly convergent with order of accuracy p. As a example, EM has the strong accuracy 0.5 and weak accuracy 1.

3.2 A Numerical Method for Sampling: The Metropolis-Hastings Algorithm

The Metropolis-Hastings Algorithm is a way to generate random variables with distributions approximating a given distribution. One important application is to calculate the integrals for expectations in high dimensions which it usually hard to use standard quadrature rules to deal with. The standard Monte Carlo method provides a convenient way to compute such integrals, but the convergence rate depends on the distribution chosen and sampling random variables with arbitrary distributions is not

simple. The Metropolis-Hastings Algorithm uses a Markov Chain to give proposal moves and then introduce random rejections with a designed probability. Then the Markov Chain with rejections will be converging to some random variable with the given distribution, therefore, after a large number of iterations, the distribution of the observations generated from the simulation is approximately the target distribution. Given a known transitional probability $q(x, y)$ for a Markov Chain X_n, from state $X_n = x$ to state $X_{n+1} = y$ with some initial state $X_0 = x_0$ and a target distribution $\pi(x)$, the MH method takes the following procedure:

- generate X^*_{n+1} from the Markov Chain with a known transitional probability $q(X_n, X^*_{n+1})$
- generate ξ from uniform distribution in $[0, 1]$
- compute

$$\alpha(X_n, X^*_{n+1}) = \begin{cases} \min\left(\frac{\pi(X^*_{n+1})q(X^*_{n+1},X_n)}{\pi(X_n)q(X_n,X^*_{n+1})}, 1\right) & \text{if } \pi(X_n)q(X_n, X^*_{n+1}) > 0 \\ 1 & \text{otherwise} \end{cases}$$

- accept X^*_{n+1} with probability $\alpha(X_n, X^*_{n+1})$

$$X_{n+1} = \begin{cases} X^*_{n+1} & \text{if } \xi < \alpha(X_n, X^*_{n+1}) \\ X_n & \text{otherwise.} \end{cases}$$

3.3 Detailed Balance and Metropolis-Hastings Algorithm

If we expect a Markov chain to converge, we need to ask first if it has a stationary distribution. Given the process defined on $\mathbb{R}^d$, we call $\pi^*(x)$ a stationary distribution for a Markov chain, if

$$\pi^*(dy) = \int_{\mathbb{R}^d} P(x, dy)\pi^*(dx)$$

where $P(x, dy)$ is the transitional distribution. A sufficient condition for the Markov chain to have a stationary distribution π^* is the detailed balance condition (in Markov processes, this is the reversibility) with respect to π^*.

Let $S_{xy} = S_x \times S_y$ where S_x and S_y are measurable sets in $\mathbb{R}^d$. Then the detailed balance condition with respect to a probability distribution π for a Markov process with the transitional probability distribution P^t is

$$\int_{S_{xy}} \pi(dx)P^t(x, dy) = \int_{S_{yx}} \pi(dx)P^t(x, dy)$$

This is describing the "microscopic (detailed) balance" of transitions that the probability of going from S_x to S_y should be the same as the probability of going from S_y to S_x. In the case where

the proposal step is given by some continuous distribution, we can assume that the composite transitional distribution $P(x, dy)$ has a continuous component and a discrete component:

$$P(x, dy) = p(x, y)dy + r(x)\delta_x(dy)$$

where $p(x, x) = 0$, $\delta_x(dy) = 1$ if $x \in dy$ and 0 otherwise. $r(x) = 1 - \int p(x, y)dy$ is the probability that the chain remains at x. Then for a Markov chain with transitional distribution $P(x, dy)$, the detailed balance condition with respect to π^* can be expressed in the form of

$$\pi^*(x)p(x, y) = \pi^*(y)p(y, x)$$

This is saying that the probability of transition from state x to y is the same as the probability of transition from state y to x. To see how the detailed balance condition implies the existence of a stationary distribution, we use a formal computation: for any measurable set A

$$\int P(x, A)\pi^*(x)dx = \int_{\mathbb{R}^d}\int_A p(x, y)\pi^*(x)dydx + \int_{\mathbb{R}^d} r(x)\delta_x(A)\pi^*(x)dx$$

$$= \int_A\int_{\mathbb{R}^d} p(y, x)\pi^*(y)dxdy + \int_A r(x)\pi^*(x)dx = \int_A (1 - r(x))\,\pi^*(y)dy + \int_A r(x)\pi^*(x)dx = \int_A \pi^*(y)dy.$$

The Metropolis-Hastings Algorithm is constructed to satisfy the detailed balance condition with respect to the target distribution π by adjusting the transitional probability. Loosely speaking, if the Markov chain that generates the proposal step satisfies

$$\pi(x)q(x, y) > \pi(y)q(y, x)$$

then jumps going from x to y occur too often to have π as a stationary distritbution. By introducing an acceptance probability of the move α, the Markov chain with rejection has transitional probability

$$\pi(x)q(x, y)\alpha(x, y) = \min\left(\pi(x)q(x, y), \pi(x)q(x, y)\right) = \pi(y)q(y, x)\alpha(y, x)$$

Therefore, the detailed balance condition with respect to π is satisfied and the chain has a stationary distribution π. A sufficient condition for the convergence of Metropolis-Hastings Algorithm from the theory of Markov processes is the aperiodicity and irreducibility of the proposal chain. This is generally not hard to achieve: for example, the proposal move from the Brownian motion satisfies the condition.

3.4 Metropolis-adjusted Langevin Algorithm

Though the convergence of Metropolis-Hastings algorithm is guaranteed, the speed of convergence may still be a problem. A slow convergence rate will hugely affect the efficiency of the sampling.

One solution to this is to construct a better Markov chain that gives proposal steps. Roberts and Tweedie considered and showed that the Langevin equation

$$dX = \frac{1}{2}\nabla \ln \rho_{eq}(X)dt + dB(t) \tag{3.2}$$

possesses a unique stationary distribution ρ_{eq} and the distribution of $X(t)$ converges to ρ_{eq} exponentially [36]. However, Roberts and Tweedie also pointed out that for naive time discretizations like the Euler-Maruyama scheme (EM) or Milstein's method, they will typically not preserve the correct equilibrium density. They also showed that these methods may not be ergodic at all even though the underlying diffusion is exponentially ergodic. Therefore, they proposed Metropolis-adjusted Langevin algorithm: an algorithm takes proposal steps by the Euler-Maruyama method and then uses rejections according to Metropolis-Hastings algorithm to preserve the exact stationary distribution.

Let h be the step length of the numerical discretisation, the iteration of MALA takes the following procedure:

$$X^*_{n+1} = X_n + \frac{1}{2}\nabla \ln \rho_{eq}(X_n)h + B((n+1)h) - B(nh) \tag{3.3}$$

where B is standard d-dimensional Brownian motion. X_{n+1} is given by

$$X_{n+1} = \begin{cases} X^*_{n+1} & \text{if } \xi_n < \alpha_h(X_n, X^*_{n+1}), \\ X_n & \text{otherwise,} \end{cases} \tag{3.4}$$

where

$$\alpha_h(x,y) = \min\left(1, \frac{q_h(y,x)\rho_{\mathsf{eq}}(y)}{q_h(x,y)\rho_{\mathsf{eq}}(x)}\right).$$

and $\xi_n, n \geq 1$ is an independent, identically distributed sequence of random variables, uniform on $[0,1]$ and independent of B. Here

$$q_h(x,y) = \frac{1}{\sqrt{2\pi h}} \exp\left(-\left(x + \frac{1}{2}\nabla \ln \rho_{eq}(x)h - y\right)^2 / 2h\right)$$

is the transition probability density for X^*_{n+1} being at y given that X_n is at x. Since Roberts and Tweedie were only interested in sampling the equilibrium distribution, they did not analyze the convergence of the Metropolized scheme as a numerical integrator. Bou-Rabee and Vanden-Eijnden [5] showed that MALA as an numerical integrator for SDEs is not only ergodic with respect to ρ_{eq} but also converges to the solution of the SDE strongly.

3.5 Our numerical method

If D and ρ_{eq} are smooth, then MALA is both a convergence method and a method that preservers the given stationary distribution for the SDE (2.6). However, in the model problem in 2.1, both D and

ρ_{eq} have discontinuities in which case the drift term in (2.5) has a singularity and needs to be defined as a distribution. On the other hand, for SDEs that have variable diffusion coefficients and satisfy the detailed balance condition with respect to a known equilibrium distribution, in 2.3 we proposed a framework for such systems in which, instead of a diffusion coefficient and a drift coefficient, a modeller specifies a diffusion coefficient and a equilibrium distribution, and then assumes detailed balance with respect to this equilibrium distribution. Therefore, it is possible to work directly with the diffusion coefficient and equilibrium distribution, rather than the drift coefficient.

Our method is a variant of the MALA scheme. Instead of using a convergent scheme for the SDE, we only use the diffusion coefficient to give a trial step and then use the Metropolis-Hastings rejection procedure to guarantee the correct equilibrium density. Therefore the drift is enforced only indirectly through the rejection step. The motivation for this idea is that for any SDE the infinitesimal drift is uniquely determined by the infinitesimal diffusion, the equilibrium distribution, and the detailed balance condition [43]. Therefore, if we have a Markov chain that approximates a diffusion process with the correct diffusion coefficient and the correct equilibrium distribution, and also satisfies the detailed balance condition, we expect that the process also has approximately the correct drift coefficient.

For our scheme, since the trial step is given with the correct diffusion and the Metropolis-Hastings rejection process provides the detailed balance with respect to the correct equilibrium density, we expect that it converges to the correct solution to the stochastic differential equation. A similar theorem appears in [4] for general self-adjoint diffusions (where the density is not necessarily integrable). In this chapter, we will show directly that the process has the correct drift and diffusion in the limit of steplength going to zero, when the coefficients are sufficiently smooth. In particular, we show that the scheme is weakly convergent with order of accuracy $1/2$ under appropriate conditions.

Though our results here are for smooth coefficients, the main motivation for our scheme is to to handle instances of (2.6) where D has jump discontinuities. Other work has developed numerical schemes for similar classes of problems. The reference [22] proposes a method for such systems that does not make explicit use of the equilibrium distribution and hence does not preserve it exactly. However, their method could be adjusted with a Metropolis-Hastings step in order to do so. Another approach is to resolve the jump discontinuities in D by developing a separate procedure for when the state of the system approaches the discontinuity. This approach is taken by [13, 25, 28] for one-dimensional systems, who make use of the theory of skew Brownian motion to resolve the discontinuity.

Here we define our algorithm from [43] for approximating the solution of (2.6). Let h be the step length. The trial step is given by

$$X^*_{n+1} = X_n + \sqrt{2D(X_n)}[B((n+1)h) - B(nh)]. \tag{3.5}$$

This is accepted with probability α_h

$$X_{n+1} = \begin{cases} X^*_{n+1}, & \text{if } \xi_n < \alpha_h\left(X_n, X^*_{n+1}\right), \\ X_n, & \text{otherwise.} \end{cases} \tag{3.6}$$

where ξ_k satisfies uniform distribution on [0,1] and α_h is the acceptance probability for Metropolis-Hastings rejection procedure [36] from state X_n to X^*_{n+1} with the expression

$$\alpha_h(x,y) = \min\left(1, \frac{q_h(y,x)\rho_{eq}(y)}{q_h(x,y)\rho_{eq}(x)}\right) \tag{3.7}$$

and $q_h(x,y)$ is the transitional probability density determining the trial step (3.5)

$$q_h(x,y) = \frac{1}{(4\pi h D(x))^{d/2}} \exp\left(-\frac{(x-y)^2}{4hD(x)}\right). \tag{3.8}$$

This choice of α_h and q_h in the Metropolis-Hastings rejection process guarantees that the process $X_n, n = 0, 1, 2, \ldots$ satisfies detailed balance with respect to the density ρ_{eq}. The Metropolis rejection procedure guarantees that the process $X_n, n \geq 1$ has $\rho_{\mathsf{eq}}(x)$ as its equilibrium density. On the other hand, in any region of the state space where $D(x)$ and $\rho_{\mathsf{eq}}(x)$ are constant with respect to x, α_h is 1 and so the method reduces to the Euler-Maruyama method for the constant coefficient diffusion without drift. Notice that the algorithm only requires the ratio of the equilibrium density at different positions. The method may also be applied to problems with non-normalizable equilibrium distribution (i.e. the equilibrium distribution is no more a probability distribution).

3.6 Numerical Simulations

In this section we validate our method with the following numerical experiments. We chose 1-dimensional examples of (2.6) with the following features:

1. Smooth diffusion coefficient D and equilibrium density ρ_{eq}, for which we have an exact solution.
2. Smooth and periodic diffusion coefficient D and equilibrium distribution $\rho_{eq} = 1$ which is not normalizable.
3. Geometric brownian motion, for which we have a non-normalizable ρ_{eq} and degenerate D.
4. Piecewise constant D and uniform ρ_{eq} in $[-1, 1]$.

The motivation of these examples is to demonstrate the convergence of the numerical scheme for some problems not necessarily satisfying the conditions of Theorem 3.

3.6.1 Example 1: SDE with smooth coefficients.

We first test the method on a SDE for which we have a closed-form solution,

$$dX = -\frac{X}{2}dt + \sqrt{1-X^2}dB. \tag{3.9}$$

Comparing with (2.6), we can see that this is the case when the diffusion coefficient is

$$D(x) = \frac{1-x^2}{2}$$

and the equilibrium density is

$$\rho_{eq}(x) = \frac{1}{\pi\sqrt{1-x^2}}$$

in the domain $|x| < 1$. If the initial condition is $X(0) = \frac{1}{2}$, then it has the exact solution

$$X(t) = \sin(B(t) + \frac{\pi}{6}).$$

The equation (3.9) does not satisfy the conditions of Theorem 3 because D is not bounded away from zero and $\frac{d}{dx}\ln(\rho_{eq}(x))$ approaches infinity at $x = \pm 1$.
Firstly, we numerically verify that this method keeps the exact equilibrium density ρ_{eq} and approximates the given diffusion coefficient. To compute these statistical quantities of the trajectories, the domain $(-1, 1)$ is cut into 20 equally spaced subintervals $[x_i, x_{i+1}), i = 0, \ldots, 19$. The density is computed by dividing the number of times that the particle is in the particular interval over the total number of time steps. The effective diffusion coefficient is computed as in [43]:

$$D(x_i) = \underset{X_{kh}\in[x_i,x_{i+1})}{\text{mean}} \frac{(X_{(k+1)h} - X_{kh})^2}{2h}$$

The SDE is simulated with different time step lengths over a total time interval of length $T = 1000$. With these parameters we plot the values of $\rho_{eq}(x)$ and $D(x)$ over the domain $|x| < 1$ in Figure 3.1. The error bars show estimates of standard error due to the finite time simulation. As we can see from Figure 3.1, the numerical method produces the correct distribution for all the time step lengths, while the effective D is converging to the exact curve as the time step length is decreasing.

In order to check the weak accuracy of the numerical scheme, we measure the mean error at time $T = 1$ with test function $f(x)$ as in [18],

$$\epsilon_h = |\mathbb{E}(f(X_{Nh})) - \mathbb{E}(f(X(T)))| \tag{3.10}$$

The expectation $\mathbb{E}(f(X_{Nh}))$ is approximated by the average values of $f(X_{Nh})$ over a number of $M = 10^7$ trajectories. Figure 3.2 shows the error versus the time step length with test functions $f(x) = x$ and $f(x) = x^2$. For these test functions, the exact solutions are $\mathbb{E}X(1) = \frac{1}{2\sqrt{e}}$, $\mathbb{E}(X(1))^2 = \frac{1}{2} - \frac{1}{4e^2}$ The plot shows the accuracy is of order $\frac{1}{2}$.

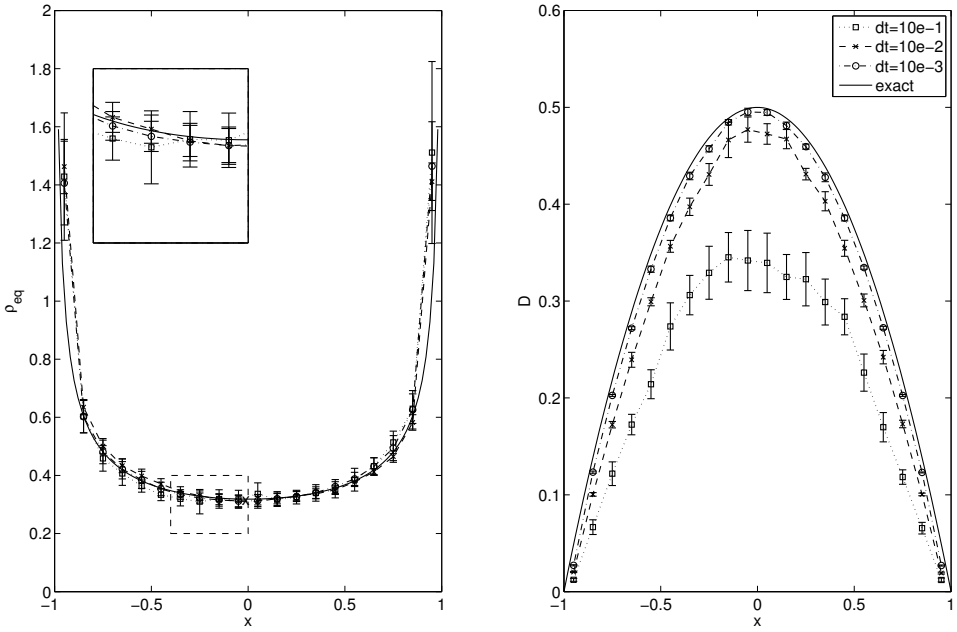

Figure 3.1: Plot of the computed equilibrium density (left) and diffusion coefficient (right) for Example 1.

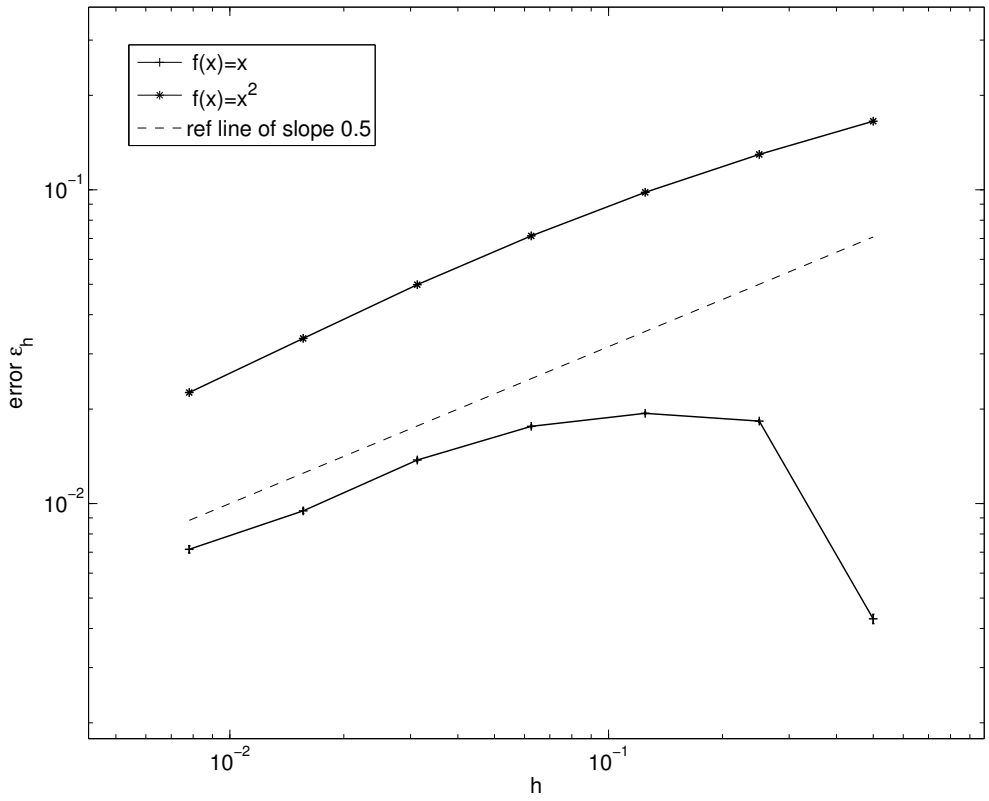

Figure 3.2: The weak error of the scheme for Example 1 with test function $f(x) = x$ and $f(x) = x^2$. The error bars in this plot are smaller than the symbol size.

3.6.2 Example 2: SDE with smooth coefficients.

Here we consider the case with smooth diffusion coefficient $D = \sin(x) + 2$ and uniform equilibrium distribution $\rho_{eq} = 1$. Using (2.6), this gives the SDE

$$dX(t) = \cos(X)dt + \sqrt{4 + 2\sin(x)}dB$$

with initial condition $X(0) = 0$. Here, ρ_{eq} is not normalizable since the domain is infinite, therefore we do not have a probability density at equilibrium. However, computationally, since we only simulate to finite time, we can still look at the probability distribution of $X(T)$ and its expectation and moments are well defined. For this SDE, since we do not have the exact solution, we measure the error by subtracting the results from time step length $h/2$ from h, i.e.

$$\epsilon_h = \left|\mathbb{E}(f(X_h(T))) - \mathbb{E}(f(X_{h/2}(T)))\right| \tag{3.11}$$

The expectation is approximated by the average over $M = 4 \times 10^6$ trajectories. Figure 3.3 shows the error plot compared with the error from Euler-Maruyama (EM) scheme. The EM method shows the expected weak accuracy of order 1. Our method shows the weak accuracy of order $\frac{1}{2}$ for the test function $f(x) = x^2$. Furthermore, we observe super-convergence with apparent order 1 for test function $f(x) = x$. A closer look at the leading $\sqrt{h}$ term in the error shows that its coefficient in this case is comparably smaller than the next term due to the effect of $f(x)$ being odd. Therefore, when h is not small enough, the error is dominated by the order h term.

3.6.3 Example 3: Geometric Brownian Motion.

For this example, we test our scheme on geometric brownian motion

$$dX(t) = aXdt + bXdB$$

with $a = 1$, $b = 1$ are constants. The initial condition is $X_0 = 1$. We have the exact solution

$$X(t) = X_0 \exp\left(\left(a - \frac{b^2}{2}\right)t + bB(t)\right) = X_0 \exp\left(\frac{1}{2}t + B(t)\right)$$

with expectation

$$\mathbb{E}(X(t)) = X_0 \exp(t).$$

Firstly we need to rewrite the equation in the form of (2.6). Notice that even though geometric brownian motion does not have an equilibrium density, we can still formally let

$$D = \frac{1}{2}X^2, \;\; \rho_{eq} = 1$$

to get the same form of SDE as we want. Figure 3.4 shows the weak error with test function $f(x) = x$ at time $t = T = 1$ compared with the error from the Euler-Maruyama scheme. The

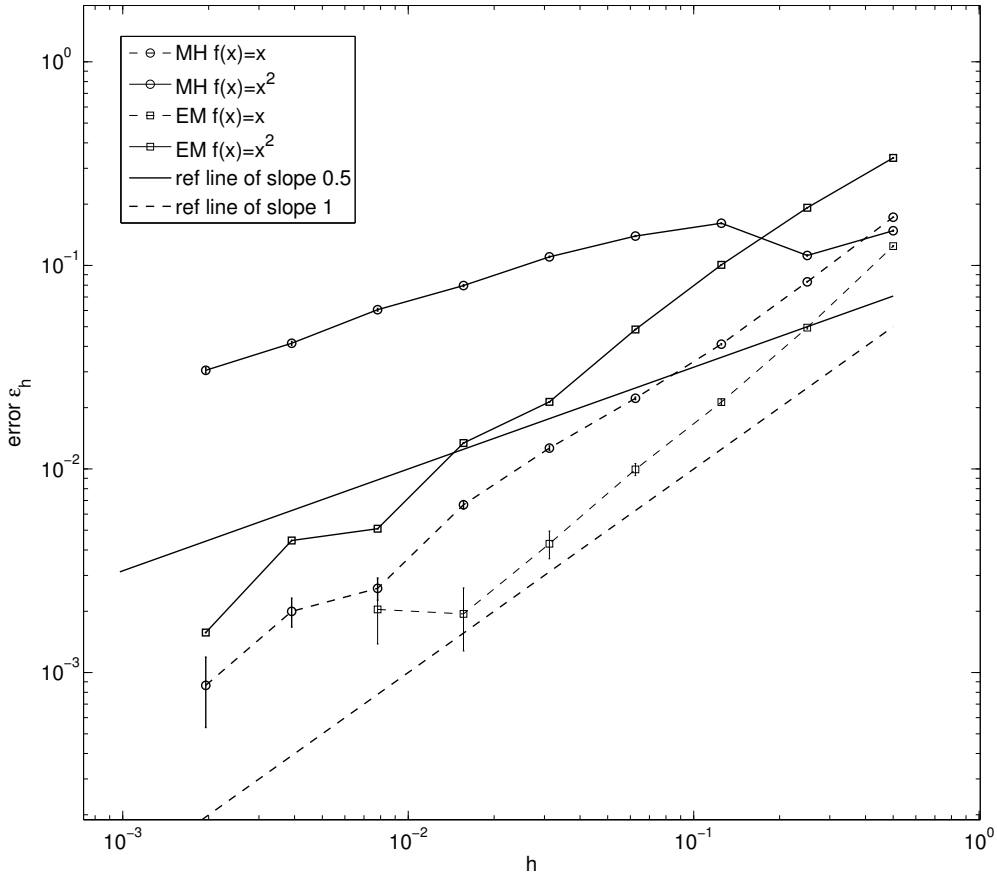

Figure 3.3: The weak error of the numerical schemes: Euler-Maruyama (EM) and our scheme (MH) in Example 2 with test functions $f(x) = x, x^2$

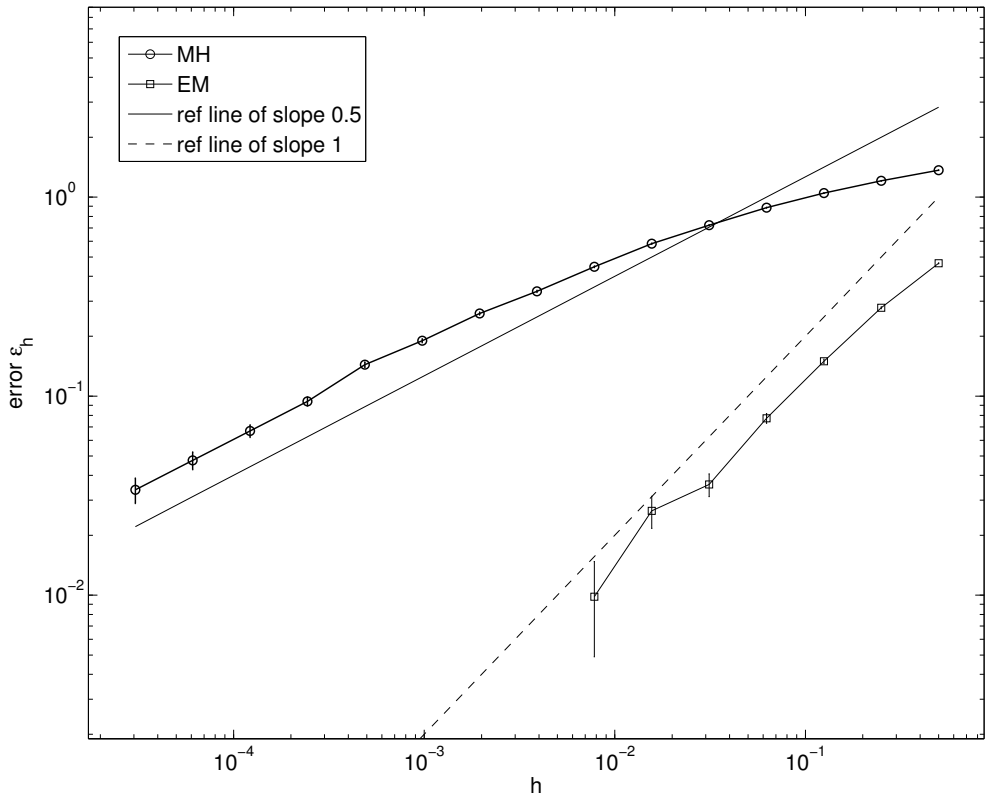

Figure 3.4: The weak error of the numerical schemes: Euler-Maruyama (EM) and our scheme (MH) in Example 3 with test functions $f(x) = x$. Error is measured using (3.10).

error is measured over $M = 5 \times 10^5$ trajectories, using (3.10) and (3.11). As a result, though geometric brownian motion does not satisfy the conditions in Theorem 3, the numerical simulation still demonstrates that we can expect convergence in this case with weak accuracy of order $\frac{1}{2}$.

3.6.4 Example 4: SDE with piecewise constant diffusion coefficient.

Here we numerically demonstrate the convergence of the method for the simple case where

$$D(x) = \begin{cases} 1 & \text{if } x \leq 0, \\ 2 & \text{if } x > 0, \end{cases}$$

and

$$\rho_{\mathsf{eq}}(x) = \begin{cases} 1 & \text{if } x \in [-1, 1], \\ 0 & \text{otherwise.} \end{cases}$$

Equation (2.6) with this choice of $D(x)$ and $\rho_{\mathsf{eq}}(x)$ models a particle diffusing on the interval $[-1, 1]$ with reflecting boundary conditions at ± 1 and a piecewise constant diffusion coefficient. The reflecting boundary conditions are conveniently implemented by our choice of $\rho_{\mathsf{eq}}(x)$.
Figure 3.5 shows the results of simulating (2.6) with these choices of $D(x)$ and $\rho_{\mathsf{eq}}(x)$ using the method we have described. To show results, we divide the interval $[-1, 1]$ into 20 equal subintervals, and plot the density for the amount of time the particle spends in each subinterval. We also plot the effective diffusion coefficient for each bin, which we define to be the average observed value of $(X_{n+1} - X_n)^2/2h$ over the trajectory, for all n such that X_n lies in the given bin. Simulations were conducted with $h = 0.01, 0.001, 0.0001$ and for trajectories long enough so that standard statistical errors in the plots are smaller than the symbols used. We see that for all values of h the equilibrium density $\rho_{\mathsf{eq}}(x)$ is correctly reproduced. The effective diffusion coefficient converges to D as h goes to zero.

The weak error in this example is calculated using formula

$$\epsilon_h = \left| \mathbb{E}(f(X_h(T))) - \int_{x \in \mathbb{R}} f(x)\rho(x, T)dx \right|$$

where $\rho(x, t)$ solves the corresponding Fokker-Plank equation,

$$\frac{\partial \rho(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(D(x)\frac{\partial}{\partial x}\rho(x,t) \right)$$

with homogeneous Neumann boundary conditions $\frac{\partial}{\partial x}\rho_{eq}(x, t) = 0$ at $x = \pm 1$ and initial condition $\rho(x, 0) = \delta(x)$ where $\delta(x)$ is the delta distribution. This divergence form PDE is solved numerically using Crank-Nicolson(CN) scheme with a very fine mesh. The expectation is approximated by averaging over $M = 4 \times 10^6$ trajectories. Figure 3.6 shows the convegence of the method with test functions $f(x) = x$ and $f(x) = x^2$. In each case we see order $\frac{1}{2}$ convergence despite the discontinuity of D at $x = 0$.

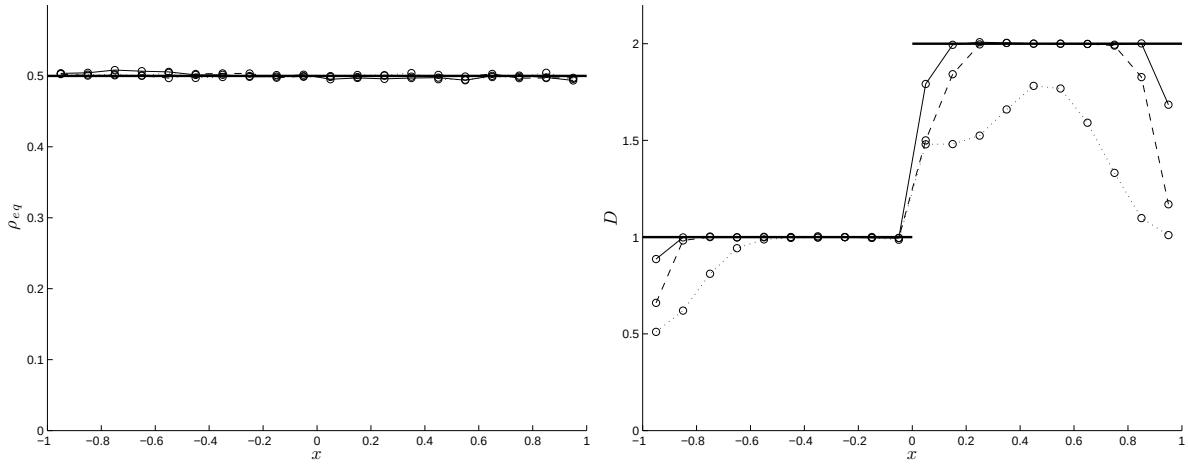

Figure 3.5: Equilibrium density ρ_{eq} and local diffusion coefficient D for method (3.5) applied to a simple one-dimensional SDE. Shown are results for $h = 0.01$ (dotted), $h = 0.001$ (dashed), and $h = 0.0001$ (solid), along with the exact values for the SDE.

3.7 Weak convergence of the method

Firstly, we exhibit some sufficient conditions on D and ρ_{eq} for the ergodicity of the SDE (2.6) and the numerical scheme in Theorem 1 and Theorem 2. The convergence to the equilibrium ρ_{eq} of (2.7) is shown using the idea of relative entropy and the logarithmic Sobolev inequality [1]. As we show in Theorem 2, the numerical method is ergodic and has the correct equilibrium distribution because of the use of the Metropolis-Hastings method. We then show in Theorem 3 that the numerical method converges weakly with order $1/2$ for smooth ρ_{eq} and D.

We will let $H(\rho_1|\rho_2)$ be the relative entropy of ρ_1 with respect to ρ_2 where

$$H(\rho_1|\rho_2) := \int_{\mathbb{R}^d} \rho_1(x) \ln \frac{\rho_1(x)}{\rho_2(x)} dx.$$

The reason to use relative entropy is due to the Csiszàr-Kullback inequality

$$H(\rho_1|\rho_2) \geq \frac{1}{2} \|\rho_1 - \rho_2\|^2_{L^1}. \tag{3.12}$$

Therefore, once we have convergence in the relative entropy, we will have convergence in L^1. Another useful functional $I(\rho_1|\rho_2)$ called the entropy dissipation functional is defined by

$$I(\rho_1|\rho_2) := \int_{\mathbb{R}^d} \rho_1(x) \nabla \ln \frac{\rho_1(x)}{\rho_2(x)} \cdot \nabla \ln \frac{\rho_1(x)}{\rho_2(x)} dx.$$

Theorem 1. *Suppose*

1. *The known equilibrium probability density $\rho_{eq} \in C^2(\mathbb{R}^d)$ is positive $\rho_{eq}(x) > 0$ and satisfies $\nabla^2 \ln \rho_{eq} \leq -\lambda I_d$, where I_d is the identity matrix of dimension d and $\lambda > 0$ is some positive constant.*

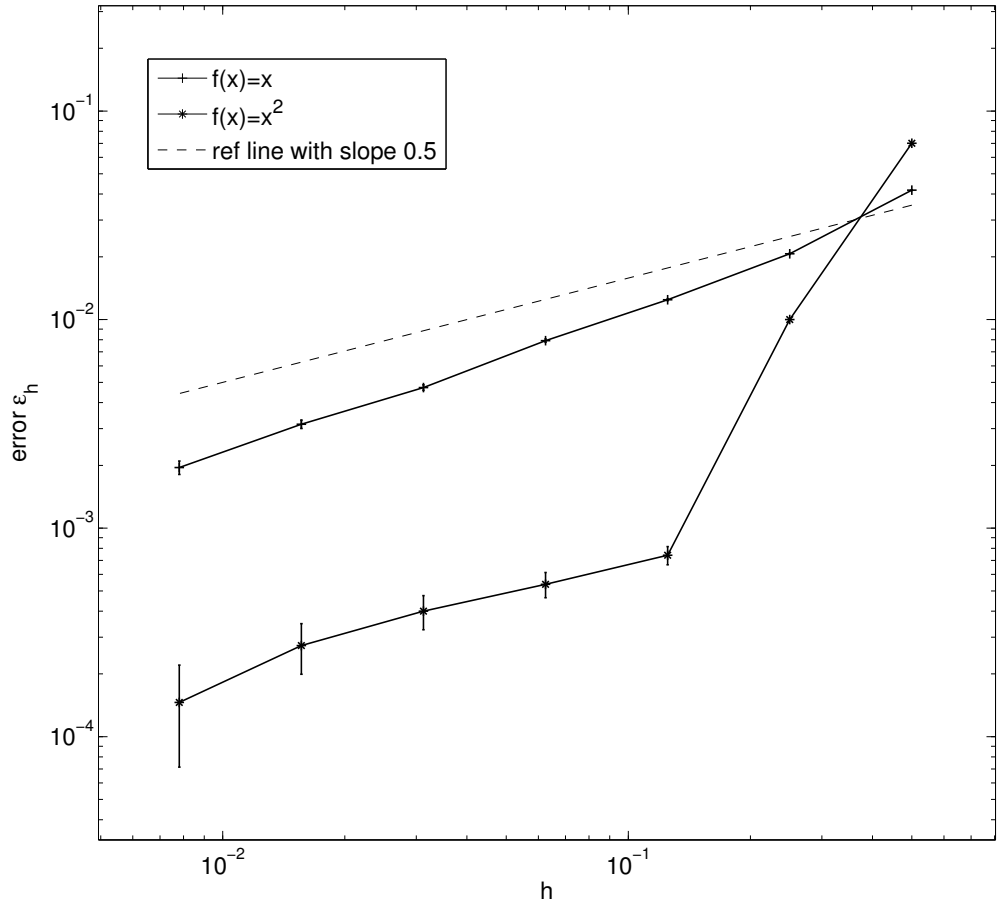

Figure 3.6: The weak error of the numerical scheme in Example 4 with test functions $f(x) = x, x^2$.

2. $H(\rho(x,0)|\rho_{eq}(x)) < \infty$. *i.e. the initial condition of* (2.7) *has finite relative entropy with respect to the equilibrium density* ρ_{eq}.

3. *The diffusion coefficient* $D \in C^2(\mathbb{R}^d)$ *and* D *is bounded below by some positive number:* $\inf D(x) = D_{\min} > 0$.

4. *The surface integral*

$$\lim_{R\to+\infty} \int_{|x|=R} D\rho_{eq} \left| \nabla \frac{\rho}{\rho_{eq}} \right| dx = 0 \tag{3.13}$$

then $\rho(x,t)$ *in* (2.7) *converges to* ρ_{eq} *exponentially fast in relative entropy.*

$$H(\rho(x,t)|\rho_{eq}) \leq e^{-2t\lambda D_{\min}} H(\rho(x,0)|\rho_{eq})$$

Hence, $\rho(x,t) \to \rho_{eq}(x)$ *in* L^1 *as* $t \to \infty$.

Proof. Let $g = \frac{\rho}{\rho_{eq}}$. Assuming ρ is a solution to (2.7), g will satisfy

$$\frac{\partial g}{\partial t} = \frac{\nabla \cdot (D\rho_{eq} \nabla g)}{\rho_{eq}}.$$

Let $\phi(g) = g \ln g - g + 1$, then through direct calculation, $H(\rho|\rho_{eq}) = \int_{\mathbb{R}^d} \phi(g)\rho_{eq} dx$ and

$$\frac{d}{dt} H(\rho|\rho_{eq}) = \int_{\mathbb{R}^d} \frac{\partial \phi(g)}{\partial g} \frac{\partial g}{\partial t} \rho_{eq} dx = -\int_{\mathbb{R}^d} D\rho \left(\nabla \ln \frac{\rho}{\rho_{eq}} \cdot \nabla \ln \frac{\rho}{\rho_{eq}} \right) dx \leq -D_{\min} \cdot I(\rho(x,t)|\rho_{eq}(x))$$

where the surface integral from integration by parts vanishes because of (3.13). By Theorem 1 in [27] (See Theorem 9 in Appendix B), Condition 1 here guarantees that the logarithmic Sobolev inequality with parameter λ holds

$$H(\rho|\rho_{eq}) \leq \frac{1}{2\lambda} I(\rho|\rho_{eq}).$$

As a result,

$$\frac{d}{dt} H(\rho|\rho_{eq}) \leq -2\lambda D_{\min} H(\rho|\rho_{eq})$$

We get the exponential convergence in relative entropy which will imply exponential convergence in L^1 by (3.12). □

Remark: Theorem 1 also works when ρ_{eq} is only positive in some connected open set $\mathcal{D} \subset \mathbb{R}^d$ provided that condition 4 is replaced by zero-flux boundary conditions on $\partial\mathcal{D}$. By restricting the domain inside the region, ρ_{eq} will be strictly positive inside the domain and there's no problem of dividing by zero. A discussion about relaxing the uniform convexity of $\ln \rho_{eq}$ in condition 1 can be found in [27].

Theorem 2. *Suppose the diffusion coefficient D is bounded below by some positive number: $\inf D(x) > 0$ and suppose ν is the equilibrium probability distribution with density ρ_{eq}. Let the numerical scheme defined in* (3.5)*,* (3.6) *generate a Markov chain with n-step transitional probability distribution $P^n(x,\cdot)$. Then $P^n(x,\cdot)$ converges to the equilibrium probability distribution $\nu(\cdot)$ in total variation norm as $n \to \infty$ i.e. :*

$$\sup\{|P^n(x,A) - \nu(A)| : \textit{for all measurable set } A\} \to 0 \textit{ , uniformly in x}$$

Proof. The proof follows from [20, 41] (see Theorem 10 in Appendix B). We only need to show that the chain generated by the numerical method is ρ_{eq}-irreducible and aperiodic. These two conditions are satisfied since 1) our proposal step is given by Gaussian random variables which gives a positive probability to any set with positive Lebesgue measure, 2) the acceptance rate $\alpha_h(x,y)$ in Metropolis-Hastings rejection step will always be positive as long as $\rho_{eq}(y)$ is positive. Hence the transitional distribution of the Markov chain with rejections generated by the numerical method will have a positive probability of jumping into any set where ρ_{eq} is positive. □

Now we show the main result of this chapter that the scheme is weakly convergent. We rewrite the time stepping of the scheme in the form

$$X_{n+1} = X_n + \bar{A}(X_n, h; X^*_{n+1}, \xi_n)$$

where

$$\bar{A}(X_n, h; X^*_{n+1}, \xi_n) = (X^*_{n+1} - X_n)\mathbb{1}_{\xi_n < \alpha_h\left(X_n, X^*_{n+1}\right)}$$

is the increment of the numerical scheme in a single step. We shall use A to denote the increment of the exact solution in a single step.

Theorem 3 (Weak convergence of the scheme)**.** *Let $\|\cdot\|_2$ be the 2-norm for matrix and ∇^2 be the Hessian. Suppose that*

1. *the coefficients $a(x) = \nabla D(x) + D(x)\nabla \ln \rho_{eq}(x)$ and $b(x) = \sqrt{2D(x)}$ of equation* (2.6) *are continuous, satisfy a Lipschitz condition*

   $$|a(x) - a(y)| + |b(x) - b(y)| \le K\,|x - y|$$

   and together with their partial derivatives with respect to x of order up to and including 3 have at most polynomial growth.

2. *$\|\nabla^2 D(x)\|_2$ and $\|\nabla^2 \ln \rho_{eq}(x)\|_2$ can be bounded by some polynomial in x and the diffusion coefficient $D(x)$ is bounded away from zero: $\inf(D(x)) > 0$.*

3. *the function $f(x)$ together with its partial derivatives of order up to and including 3 have at most polynomial growth.*

We assume the initial condition $X(0) = x_0$ is fixed. For uniform discretization $t_k = hk$, $k = 1, ..., N$ with $t_N = T$ the total time, the following inequality holds for all k:

$$|\mathbb{E}(f(X(t_k)) - f(X_k))| \leq Ch^{1/2}.$$

We prove Theorem 3 by analyzing the local error of the scheme. Bou-Rabee et al. [4] have shown a similar result about a more general class of Metropolized integrators applied to a class of equations including our (2.6). Their method consists of the use of a Runge-Kutta type integrator for the trial step followed by a Metropolis-Hastings decision to accept or reject the step. In general, their trial step uses the gradient of the diffusion coefficient, but also allows our choice of using only the diffusion coefficient itself as a special case (corresponding to $G_h = 0$ in their notation). In the following estimates of the local error, we use the same techniques as [4], making precise the dependence of the remainder term on x in order to guarantee global convergence. We also have a slightly less restrictive condition on $\rho_{eq}(x)$ in that derivatives of $\ln \rho_{eq}(x)$ do not need to be bounded.

Proof. We are going to apply Theorem 2.1 of [29, p. 100] (see Theorem 11 in Appendix B) to show the weak convergence of the scheme. The condition (a) of their theorem corresponds to our condition 1 which is the requirement on the smoothness and the growth of the coefficients a and b. The condition (c) there corresponds to our condition 3 which is the requirement on the smoothness and the growth of the test function f. Their condition (d) is a uniform *a priori* bound on the moments of the numerical scheme which is guaranteed by our Lemma 1. What remains to be shown is their condition (b): bounds on the moments of the increments of the numerical method. For convenience, we use $\mathcal{O}(x, h^p)$ to denote a quantity that can be bounded by $K_1(x)h^p$ where $K_1(x)$ is some polynomial or a matrix of polynomial entries.
The condition (b) in Theorem 2.1 of [29, p. 100] has two requirements. Firstly, all the third moments of the increment in the numerical scheme must be $\mathcal{O}(X_n, h^{3/2})$, i.e.

$$\mathbb{E}_{X_n}\left(\prod_{j=1}^{3} |\bar{A}^{i_j}|\right) \leq K_1(X_n)h^{3/2}$$

Here $\bar{A}^{i_j}$ is the $i_j \in \{1, .., d\}$ component of $\bar{A}$ and $K_1(x)$ is a function with at most polynomial growth. Then, the difference between the first and second moments of the approximated increment and the exact increment needs to be $\mathcal{O}(X_n, h^{3/2})$, i.e.

$$\left|\mathbb{E}_{X_n}\left(\prod_{j=1}^{s} \bar{A}^{i_j} - \prod_{j=1}^{s} A^{i_j}\right)\right| \leq K_2(X_n)h^{3/2}, \quad s = 1, 2$$

Here $K_2(x)$ is also a function with at most polynomial growth.
For the first requirement, since

$$\left|\bar{A}(X_n,h;X^*_{n+1},\xi_n)\right| = \left|(X^*_{n+1}-X_n)\mathbb{1}_{\xi_n<\alpha_h\left(X_n,X^*_{n+1}\right)}\right| \leq \left|\sqrt{2D(X_n)}\left(B^{i_j}((n+1)h)-B^{i_j}(nh)\right)\right|$$

therefore

$$\mathbb{E}_{X_n}\left(\prod_{j=1}^{3}\left|\bar{A}^{i_j}\right|\right) \leq (2D(X_n))^{3/2}h^{3/2}$$

By the Lipschitz condition on $b(x)=\sqrt{2D(x)}$, $(2D(X_n))^{3/2}$ will be bounded by some polynomial. For the second requirement, consider the solution after one time step from the initial condition. Let $A(X(0),h)=X(h)-X(0)$ be a column vector of the increment of the exact solution.

$$\mathbb{E}_{X_0}(\bar{A}(X_0,h;X^*_1,\xi_1)-A(X_0,h)) = \mathbb{E}_{X_0}((X^*_1-X_0)\alpha_h(X_0,X^*_1)-(X(h)-X_0))$$

$$\mathbb{E}_{X_0}(\bar{A}(X_0,h;X^*_1,\xi_1)\bar{A}^T(X_0,h;X^*_1)-A(X_0,h)A^T(X_0,h))$$
$$=\mathbb{E}_{X_0}((X^*_1-X_0)(X^*_1-X_0)^T\alpha_h(X_0,X^*_1)-(X(h)-X_0)(X(h)-X_0)^T)$$

By Theorem 4, we have

$$\mathbb{E}_{X_0}((X^*_1-X_0)\alpha_h(X_0,X^*_1)) = a(X_0)h+\mathcal{O}(X_0,h^{3/2})$$

$$\mathbb{E}_{X_0}(X^*_1-X_0)(X^*_1-X_0)^T\alpha_h(X_0,X^*_1) = b^2(X_0)hI_d+\mathcal{O}(X_0,h^{3/2})$$

Let $\mathcal{L}f(x)=a^T(x)\cdot\nabla_x f(x)+\frac{1}{2}b(x)\Delta_x f(x)$ be the infinitesimal generator of the Itô diffusion (2.6). By Ito-Taylor expansion [29, p. 99] , we have the expansion componentwise

$$\mathbb{E}_{X_0}(X(h)-X_0)^i = a^i(X_0)h+\mathbb{E}_{X_0}\left(h\int_0^h \mathcal{L}a^i(X(t))dt\right) \tag{3.14}$$

$$\mathbb{E}_{X_0}((X(h)-X_0)(X(h)-X_0)^T)^{ij} = b^2(X_0)hI_d^{ij}+\mathbb{E}_{X_0}\left(h\int_0^h \mathcal{L}\left(a^i(X(t))\cdot(X^j(t)-X^j(0))\right)dt\right)+$$

$$\mathbb{E}_{X_0}\left(h\int_0^h \mathcal{L}\left(a^j(X(t))\cdot(X^i(t)-X^i(0))+\frac{1}{2}b^2(X(t))\right)dt\right) \tag{3.15}$$

Since the integrands in the remainder terms in (3.14) (3.15) are combinations of products of X,a,b and their derivatives, by assumptions on their growth, the integrands can only have at most polynomial growth in X. We can find m large enough, s.t.

$$\left|\mathbb{E}_{X_0}\left(h\int_0^h \mathcal{L}a^i(X(t))dt\right)\right| < h\mathbb{E}_{X_0}\int_0^h C_1(1+|X(t)|^{2m})dt$$

for some constant C_1. Theorem 4 in [15, p. 48] (see Theorem 13 in Appendix B) shows that the moments of the solution could be uniformly bounded by the moments of the initial condition, i.e.

$$\mathbb{E}_{X_0}\int_0^h C_1(1+|X(t)|^{2m})dt \le h\mathbb{E}_{X_0}C(1+|X(0)|^{2m}) = hC(1+|X(0)|^{2m})$$

The constant C in the last inequality only depends on T, m, K. The same process applies to the remainder in (3.15). As a result, (3.14) (3.15) becomes,

$$\mathbb{E}_{X_0}(X(h)-X_0)^i = a^i(X_0)h + \mathcal{O}(X(0),h^2) \tag{3.16}$$

$$\mathbb{E}_{X_0}((X(h)-X_0)(X(h)-X_0)^T)^{ij} = b^2(X_0)hI_d^{ij} + \mathcal{O}(X(0),h^2) \tag{3.17}$$

Hence, we have the weak local error,

$$\left|E_{X_0}(\bar{A}(X_0,h;X_1^*,\xi_1) - A(X_0,h))\right| \le \mathcal{O}(X(0),h^{3/2})$$

$$\left|E_{X_0}(\bar{A}(X_0,h;X_1^*,\xi_1)\bar{A}^T(X_0,h;X_1^*) - A(X_0,h)A^T(X_0,h))\right| \le \mathcal{O}(X(0),h^{3/2})$$

Therefore, by Theorem 2.1 in [29, p. 100], the method is convergent with order of accuracy $\frac{1}{2}$. □

Lemma 1. *Suppose the assumptions in Theorem 3 are satisfied. Then for every even number $2m$ the $2m$-moment of the numerical solution $\mathbb{E}|X_k|^{2m}$ exist and are uniformly bounded with respect to $k=1,...,N$, if and only if $\mathbb{E}|X_0|^{2m}$ exists.*

Proof. The result follows from Lemma 2.2 in [29] (see Theorem 12 in Appendix B), if the magnitude of $\bar{A}$ in one step is well-behaved. By using Theorem 4, the expectation of A is of order h

$$\begin{aligned}&\left|\mathbb{E}_{X_n}\bar{A}(X_n,h;X_{n+1}^*,\xi_{n+1})\right| = \left|\mathbb{E}_{X_n}\left((X_{n+1}^*-X_n)\mathbb{1}_{\xi_n<\alpha_h(X_n,X_{n+1}^*)}\right)\right| \\ &= \left|\mathbb{E}_{X_n}\left((X_{n+1}^*-X_n)\alpha_h(X_n,X_{n+1}^*)\right)\right| \le K(1+|X_n|)h\end{aligned}$$

while $\left|\bar{A}\right|$ is of order $h^{1/2}$

$$\left|\bar{A}(X_n,h;X_{n+1}^*,\xi_{n+1})\right| \le \left|X_{n+1}^*-X_n\right| \le \left|\frac{X_{n+1}^*-X_n}{\sqrt{2D(X_n)h}}\right|\sqrt{2D(X_n)h}$$

and $\frac{X_{n+1}^*-X_n}{\sqrt{2D(X_n)h}}$ satisfies standard normal distribution hence has moments of all orders. Then by Lemma 2.2 in [29], the moments of the numerical solution at time t_k: $\mathbb{E}|X_k|^{2m}$ exist and are uniformly bounded. □

Theorem 4. *With the definitions and assumptions in Theorem 3, we have the following,*

$$\mathbb{E}_{X_0}(X_1^*-X_0)\alpha_h(X_0,X_1^*) = a(X_0)h + \mathcal{O}(X_0,h^{3/2})$$

$$\mathbb{E}_{X_0}(X_1^*-X_0)(X_1^*-X_0)^T\alpha_h(X_0,X_1^*) = b^2(X_0)hI_d + \mathcal{O}(X_0,h^{3/2})$$

Proof. For convenience, let $x = X_0$, $y = X_1^*$, and we can rewrite the conditional expectation in the integral form,

$$\mathbb{E}_{X_0}(X_1^* - X_0)\alpha_h(X_0, X_1^*) = \int_{\mathbb{R}^d}(y-x)\alpha_h(x,y)q_h(x,y)dy \tag{3.18}$$

$$\mathbb{E}_{X_0}(X_1^* - X_0)(X_1^* - X_0)^T\alpha_h(X_0, X_1^*) = \int_{\mathbb{R}^d}(y-x)(y-x)^T\alpha_h(x,y)\cdot q_h(x,y)dy \tag{3.19}$$

Introducing a change of variable, let $\epsilon = \sqrt{h}$, $y - x = \epsilon z$. Therefore the transition probability density changes into

$$q_h(x,y)dy = \frac{1}{(\sqrt{4\pi h D(x)})^d}e^{-\frac{(x-y)^2}{4hD(x)}}dy = \frac{1}{(\sqrt{4\pi D(x)})^d}e^{-\frac{z^2}{4D(x)}}dz =: q(x,z)dz$$

which is independent of ϵ. Let

$$\alpha(x,z,\epsilon) = \min\left(1, \frac{q(x+\epsilon z, z)\rho_{eq}(x+\epsilon z)}{q(x,z)\rho_{eq}(x)}\right)$$

After the change of variable, (3.18) and (3.19) become,

$$\int_{\mathbb{R}^d}(y-x)\alpha_h(x,y)\cdot q_h(x,y)dy = \epsilon\int_{\mathbb{R}^d} z\alpha(x,z,\epsilon)q(x,z)dz \tag{3.20}$$

$$\int_{\mathbb{R}^d}(y-x)(y-x)^T(\alpha_h(x,y))\cdot q_h(x,y)dy = \epsilon^2\int_{\mathbb{R}^d} zz^T(\alpha(x,z,\epsilon))q(x,z)dz \tag{3.21}$$

Let

$$\beta(x,z,\epsilon) = \min\left(1, \exp\left(\epsilon\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \epsilon\frac{\nabla_x\rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)\right)$$

be an approximation for $\alpha(x,z,\epsilon)$. The motivation of β is discussed in Lemma 2. First we study the order of the error in drift. Applying the fact that $\int_{\mathbb{R}^d} zq(x,z)dz = 0$ which follows from the symmetry of q, we obtain

$$\epsilon\int_{\mathbb{R}^d} z\alpha(x,z,\epsilon)q(x,z)dz = \epsilon\int_{\mathbb{R}^d} z(\alpha(x,z,\epsilon)-1)q(x,z)dz.$$

By Lemma 2 and Lemma 3, we can obtain

$$\epsilon\int_{\mathbb{R}^d} z(\alpha(x,z,\epsilon)-1)q(x,z)dz = \epsilon\int_{\mathbb{R}^d} z(\beta(x,z,\epsilon)-1)q(x,z)dz + \epsilon\int_{\mathbb{R}^d} z(\alpha(x,z,\epsilon)$$
$$-\beta(x,z,\epsilon))q(x,z)dz = a(x)\epsilon^2 + \mathcal{O}(x,\epsilon^3)$$

Use Lemma 2 and Lemma 3 for (3.21),

$$\epsilon^2 \int_{\mathbb{R}^d} zz^T(\alpha(x,z,\epsilon)) \cdot q(x,z)dz = \epsilon^2 \int_{\mathbb{R}^d} zz^T \cdot q(x,z)dz + \epsilon^2 \int_{\mathbb{R}^d} zz^T(\beta(x,z,\epsilon)-1) \cdot q(x,z)dz +$$
$$\epsilon^2 \int_{\mathbb{R}^d} zz^T(\alpha(x,z,\epsilon)-\beta(x,z,\epsilon))q(x,z)dz = b(x)I^d\epsilon^2 + \mathcal{O}(x,\epsilon^3)$$

Recall that $\epsilon = \sqrt{h}$, therefore we have the desired bounds for local error. □

Lemma 2 (Estimates of $\alpha(x,z,\epsilon)$ and $\beta(x,z,\epsilon)$)**.** *With previous definitions, we have the following estimates. Let $g(z) \in \mathbb{R}$ be polynomial in z, then*

$$\left| \int_{\mathbb{R}^d} g(z)(\alpha(x,z,\epsilon)-\beta(x,z,\epsilon))q(x,z)dz \right| \leq K(x)\epsilon^2$$

where $K(x)$ has polynomial growth.

Proof. Rewrite α in exponent form,

$$\alpha(x,z,\epsilon) = \min\left(1, \exp\left(\ln \frac{q(x+\epsilon z,z)\rho_{eq}(x+\epsilon z)}{q(x,z)\rho_{eq}(x)}\right)\right)$$

A Taylor expansion for the exponent about $\epsilon = 0$ gives

$$\frac{q(x+\epsilon z,z)\rho_{eq}(x+\epsilon z)}{q(x,z)\rho_{eq}(x)} = \exp\left(\epsilon \frac{\nabla_x q(x,z)}{q(x,z)} + \epsilon \frac{\nabla_x \rho_{eq}(x)}{\rho_{eq}(x)} + R(x,z,\epsilon)\right)$$

Therefore β is obtained by keeping only the leading order ϵ terms.

$$\beta(x,z,\epsilon) = \min\left(1, \exp\left(\epsilon \frac{\nabla_x q(x,z)}{q(x,z)} + \epsilon \frac{\nabla_x \rho_{eq}(x)}{\rho_{eq}(x)}\right)\right)$$

$R(x,z,\epsilon)$ is the remainder given by

$$R(x,z,\epsilon) = \int_0^\epsilon \int_0^\xi \frac{\partial^2 \ln\left(q(x+\eta z,z)\rho_{eq}(x+\eta z)\right)}{\partial \eta^2} d\eta d\xi = \int_0^\epsilon \int_0^\xi \left(-\frac{d}{2}\frac{(z^T\nabla_w^2 D(w)z)}{D(w)}\right)_{w=x+\eta z} d\eta d\xi +$$
$$\int_0^\epsilon \int_0^\xi \left(\frac{d}{2}\frac{(z^T\nabla_w D(w))^2}{D^2(w)} + \frac{z^2(z^T\nabla_w^2 D(w)z)}{4D^2(w)} - \frac{z^2}{8D^3(w)}(z^T\cdot\nabla_w D(w))^2\right)_{w=x+\eta z} d\eta d\xi$$
$$+ \int_0^\epsilon \int_0^\xi \left(\frac{z^T\cdot\nabla_w^2\rho_{eq}(w)\cdot z}{\rho_{eq}(w)} - \frac{(z^T\cdot\nabla_w\rho_{eq}(w))^2}{\rho_{eq}^2(w)}\right)_{w=x+\epsilon z} d\eta d\xi$$

Consider the function $h(x) = \min(1, \exp(x))$. Since $h(x)$ is piecewise smooth, it is not hard to see that $h(x)$ is globally Lipschitz with Lipschitz constant 1. Therefore,

$$|\alpha(x,z,\epsilon) - \beta(x,z,\epsilon)| \leq |R(x,z,\epsilon)|$$

Therefore, with the assumptions that $\inf D(x) > 0$, $\|\nabla^2 D(x)\|$ bounded by some polynomial, $\|\nabla^2 \ln \rho_{eq}(x)\|$ bounded by some polynomial, we obtain

$$|R(x,z,\epsilon)| \leq K_1(x,z)\epsilon^2$$

Here $K_1(x,z)$ is some polynomial in x,z. Furthermore, since for fixed x, $q(x,z)$ is a multivariate Gaussian, we can calculate its absolute moments [16, p. 337],

$$\int_{\mathbb{R}^d} |z|^p \, q(x,z)dz = \begin{cases} \frac{S^d}{2}(2D(x))^{p/2}(p-1)!! & \text{if } p \text{ is even} \\ \sqrt{\frac{2}{\pi}}\frac{S^d}{2}(2D(x))^{p/2}(p-1)!! & \text{if } p \text{ is odd} \end{cases}$$

where S^d is the surface area of the unit hypersphere in $\mathbb{R}^d$. Since $b(x) = \sqrt{2D(x)}$ has at most polynomial growth, therefore,

$$\left| \int_{\mathbb{R}^d} g(z)(\alpha(x,z,\epsilon) - \beta(x,z,\epsilon))q(x,z)dz \right| \leq K(x)\epsilon^2$$

For $K(x)$ has at most polynomial growth. □

Lemma 3. *With previous definitions,*

$$\left| \int_{\mathbb{R}^d} zq(x,z)(\beta(x,z,\epsilon) - 1)dz - a(x)\epsilon \right| \leq K(x)\epsilon^2 \tag{3.22}$$

$$\left| \int_{\mathbb{R}^d} g(z)q(x,z)(\beta(x,z,\epsilon) - 1)dz \right| \leq K^*(x)\epsilon \tag{3.23}$$

where $K(x)$,$K^*(x)$ are polynomials in x

Proof. Since

$$\begin{aligned} \beta(x,z,\epsilon) - 1 &= \min\left(0, \exp\left(\epsilon\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \epsilon\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right) - 1\right) \\ &= \left(\exp\left(\epsilon\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \epsilon\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right) - 1\right)\mathbb{1}_{\Omega_{(x,0)}} \end{aligned}$$

Here the region is defined by,

$$\begin{aligned} \Omega_{(x,0)} &= \left\{ z \,\middle|\, \left(\exp\left(\epsilon\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \epsilon\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right) - 1\right) < 0 \right\} \\ &= \left\{ z \,\middle|\, \left(\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right) < 0 \right\} \end{aligned}$$

Therefore we can expand $\beta(x,z,\epsilon) - 1$ in the integrand in domain $\Omega_{(x,0)}$ about $\epsilon = 0$

$$\begin{aligned} &\int_{\mathbb{R}^d} zq(x,z)(\beta(x,z,\epsilon) - 1)dz = \int_{\Omega_{(x,0)}} zq(x,z)\left(\exp\left(\epsilon\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \epsilon\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right) - 1\right)dz \\ &= \int_{\Omega_{(x,0)}} zq(x,z)\left(\epsilon\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \epsilon\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)} + \epsilon^2 R(x,z,\xi(\epsilon))\right)dz \end{aligned}$$

where $\epsilon^2 R$ is the remainder given by,

$$R(x,z,\xi(\epsilon)) = \exp\left(\xi\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \xi\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)\left(\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)^2$$

with $0 < \xi(\epsilon) < \epsilon$. Notice that

$$zq(x,z)\left(\epsilon\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \epsilon\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)$$

is an even function in z. On the other hand, the integral domain $\Omega_{(x,0)}$ satisfies $\Omega_{(x,0)} \cup (-\Omega_{(x,0)}) = \mathbb{R}^{\ltimes}$. Hence the integral without $\epsilon^2 R$ term becomes

$$\begin{aligned}
&\int_{\Omega_{(x,0)}} zq(x,z)\left(\epsilon\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \epsilon\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)dz \\
&= \frac{\epsilon}{2}\int_{\mathbb{R}^d} zq(x,z)\left(\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)dz \\
&= \frac{\epsilon}{2}\int_{\mathbb{R}^d} z(\nabla \ln(\rho_{eq}(x))\cdot z - \frac{d}{2}\nabla\ln(D(x))\cdot z - \frac{z^T z}{4}\nabla(\frac{1}{D(x)})\cdot z)q(x,z)dz \\
&= (\nabla D(x) + D(x)\nabla \ln \rho_{eq}(x))\,\epsilon = a(x)\epsilon
\end{aligned}$$

Then we need to show the remainder term is indeed of order ϵ^2. Since in the domain $\Omega_{(x,0)}$, $\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)} < 0$, therefore $\exp\left(\xi\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \xi\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right) < 1$.

$$\begin{aligned}
&\left|\int_{\Omega_{(x,0)}} zq(x,z)R(x,z,\xi(\epsilon))dz\right| \\
&= \left|\int_{\Omega_{(x,0)}} zq(x,z)\left(\exp\left(\xi\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \xi\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)\left(\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)^2\right)dz\right| \\
&\le \int_{\Omega_{(x,0)}} |z|\, q(x,z)\left(\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)^2 dz \\
&\le \int_{\mathbb{R}^d} |z|\, q(x,z)\left(\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)^2 dz
\end{aligned}$$

As shown in Lemma 2, the term $\left|\frac{\nabla_x q(x,z)\cdot z}{q(x,z)} + \frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right|$ could be bounded by a polynomial $K_1(x,z)$ and since the integral $\int_{\mathbb{R}^d} |z|^p\, q(x,z)dz$ could be bounded by a polynomial $K_2(x)$, therefore there exists a polynomial $K(x)$

$$\left|\int_{\Omega_{(x,0)}} zq(x,z)R(x,z,\xi(\epsilon))dz\right| \le K(x)$$

A similar proof works for the other inequality (3.23). By Taylor expansion,

$$\begin{aligned}
&\left|\int_{\mathbb{R}^d} g(z)q(x,z)(\beta(x,z,\epsilon)-1)dz\right| \\
&= \epsilon\left|\int_{\Omega_{(x,0)}} g(z)q(x,z)\exp\left(\xi\frac{\nabla_x q(x,z)\cdot z}{q(x,z)}+\xi\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)\left(\frac{\nabla_x q(x,z)\cdot z}{q(x,z)}+\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right)dz\right| \\
&\leq \epsilon\int_{\mathbb{R}^d}|g(z)|\,q(x,z)\left|\frac{\nabla_x q(x,z)\cdot z}{q(x,z)}+\frac{\nabla_x \rho_{eq}(x)\cdot z}{\rho_{eq}(x)}\right|dz \leq K^*(x)\epsilon
\end{aligned}$$

which concludes the proof.

□

Chapter 4

Diffusion Approximations

Theorem 3 shows the weak convergence of the numerical method (3.5),(3.6) by direct calculations of local errors. This is of course not the unique way to show the convergence. In this chapter we will show the weak convergence of our method using the idea of diffusion approximations from a more probabilistic point of view. This idea provides us a more intuitive justification of the convergence of the method by showing that 1) the numerical method generates a Markov chain that is approximating some diffusion process, 2) the rejection from Metropolis-Hastings algorithm does not change the diffusion term, 3) the correct drift is guaranteed since the process generated by the numerical method satisfies the detailed balance condition. Though the rate of convergence can not be obtained in this way, the advantage of the proof is that the conditions on the smoothness of the coefficients in Theorem 3 can be relaxed. Though we show the convergence for coefficients with some smoothness (see Theorem 7), it is believed that the diffusion approximations may be a better way to obtain results for the discontinuous coefficients cases.

4.1 Diffusions and The Martingale Problem

Since the Itô integral is a martingale [30] (see A.3 for the definition), the solution to a (Ito-)SDE is also associated with a martingale. Firstly, we define the generator of an Itô diffusion [30, p. 117].

Definition 1. *Let X_t be an Itô diffusion in $\mathbb{R}^d$. The (infinitesimal) generator A of X_t is defined by*

$$Af(x) = \lim_{t \to 0^+} \frac{\mathbb{E}^x f(X_t) - f(x)}{t}; \quad x \in \mathbb{R}^d$$

Here the conditional expectation is taken on $X_0 = x$ and all the functions f such that the limit exists for all x form the domain of the operator $\mathcal{D}_A$,

Then the following theorem associates an Itô diffusion to the corresponding martingale [30, p. 140].

Theorem 5. *If X_t is an Itô diffusion in $\mathbb{R}^d$ with generator A, then for all twice continuously differentiable functions with compact support: $f \in C_0^2(\mathbb{R}^d)$, the process*

$$M_t = f(X_t) - \int_0^t Af(X_r)dr$$

is a martingale with respect to $\mathcal{M}_t$ the filtration generated by M_t

To simplify the concepts, we can identify each $\omega \in \Omega$ in the sample space with the trajectory mapping

$$\omega_t = \omega(t) = X_t^x(\omega)$$

Therefore the underlying probability space $(\Omega, \mathcal{M}, \mathbb{P}^x)$ is identified with $((\mathbb{R}^d)^{[0,\infty)}, \mathcal{B}, \tilde{\mathbb{P}}^x)$ where $\mathcal{B}$ is the Borel σ-algebra on $(\mathbb{R}^d)^{[0,\infty)}$. In fact, the converse of the theorem is also true in the sense the following definition.

Definition 2. *Let L be a semi-elliptic differential operator with its graph A of the form*

$$A = \left\{ \left(f, Lf = \sum_i a_i \frac{\partial f}{\partial x_i} + \sum_{i,j} b_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \right) \Big| f \in C_c^\infty(\mathbb{R}^d) \right\}$$

where the coefficients a_i and b_{ij} are locally bounded Borel measurable functions on $\mathbb{R}^d$. Given an initial distribution μ for X_0 on $\mathbb{R}^d$, we say that a stochastic process X_t on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ solves the martingale problem (A, μ) with respect to a filtration $\mathcal{G}_t$, if the process

$$M_t = f(X_t) - \int_0^t g(X_r)dr,$$

is a $\mathbb{P}$-martingale with respect to $\mathcal{G}_t$, for all $(f, g) \in A$. And the martingale problem is called well-posed if there is a unique process X_t in the sense of finite dimensional distributions that solves the martingale problem (A, μ). If this is true for all initial probability distributions μ, we say that the martingale problem for A is well-posed.

Stroock and Varadhan [42] showed that the SDE

$$dX_t = a(X_t)dt + b(X_t)dB_t$$

has a unique weak solution X_t if and only if the martingale problem for the graph of L

$$L = \sum_i a_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j} (bb^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$$

is well posed. Hence, the existence and uniqueness of the weak solution to the corresponding SDE has been formulated in terms of the well-posedness of the martingale problem.

4.2 Approximating Diffusions

The following theorem from [12, p. 355] to show the convergence of Markov chains to an Itô diffusion.

Theorem 6. *Let $a : \mathbb{R}^d \to \mathbb{R}^d$ be continuous and let b^2 be a continuous, symmetric, nonnegative definite, $d \times d$ matrix-valued function on $\mathbb{R}^d$. Let*

$$A = \{(f, \mathcal{L}f := \sum a_j \partial_j f + \frac{1}{2} \sum b^2_{j,k} \partial_j \partial_k f) : f \in C_0^\infty(\mathbb{R}^d)\}$$

and suppose that the martingale problem for (A, μ) has a unique solution for each $\mu \in \mathcal{P}(\mathbb{R}^d)$. Let $p_h(x, dy)$ be the transitional distribution on $\mathbb{R}^d$, and set

$$a_h(x) = \frac{1}{h} \int_{|y-x|\leq 1} (y - x) p_h(x, dy)$$

$$b^2_h(x) = \frac{1}{h} \int_{|y-x|\leq 1} (y - x)(y - x)^T p_h(x, dy)$$

Suppose for each $r > 0$ and $\xi > 0$

$$\sup_{|x|\leq r} |a_h(x) - a(x)| \to 0 \tag{4.1}$$

$$\sup_{|x|\leq r} \left|b^2_h(x) - b^2(x)\right| \to 0 \tag{4.2}$$

and

$$\sup_{|x|\leq r} \frac{1}{h} p_h(x, \{y : |y - x| \geq \xi\}) \to 0 \tag{4.3}$$

Let Y^h_n be a Markov chain with transition function p_h and define $X^h(t) = Y^h_{[nt]}$. If the distribution of $Y^h_0 \Rightarrow \mu$, then $X^h(t)$ converges in distribution to the solution of the martingale problem for (A, μ).

We use Theorem 6 to show the main result of this chapter, Theorem 7: the weak convergence of the numerical method (3.5),(3.6).
The proof takes two steps:

1. the Markov chains generated by the process (3.5),(3.6) are converging to an Itô diffusion with the same diffusion coefficient as the SDE (2.6),
2. the limiting process satisfies the detailed balance condition with respect to the known equilibrium density ρ_{eq}.

Therefore, the limiting process will have the correct drift by the discussion in Section 2.3.

Theorem 7. *Suppose*

- $b(x) = \sqrt{2D(x)}$ *is Lipschitz in* $|x| \leq r$ *with constant* L_1*,*
- $\nabla \ln \rho_{eq}(x)$ *is Lipschitz in* $|x| \leq r$ *with constant* L_2*,*
- $\nabla D(x)$ *is Lipschitz in* $|x| \leq r$ *with constant* L_3 *and* $D(x)$ *is bounded away from* 0*:* $\inf D(x) = K > 0$*.*

Then the numerical method (3.5),(3.6) *generates a process* $X_h(t)$ *converging weakly to the solution of the SDE* (2.6)*.*

Proof. We use the same change of variables as in (3.20) and (3.21). Let $h = \epsilon^2$ and $y - x = \epsilon z$. Then

$$a_h(x) = \frac{1}{h}\int_{|y-x|\leq 1}(y-x)p_h(x,dy) = \frac{1}{\epsilon}\int_{|z|\leq \frac{r}{\epsilon}} z\alpha(x,z,\epsilon)q(x,z)dz$$

$$b_h^2(x) = \frac{1}{h}\int_{|y-x|\leq 1}(y-x)(y-x)^T p_h(x,dy) = \int_{|z|\leq \frac{r}{\epsilon}} zz^T\alpha(x,z,\epsilon)q(x,z)dz$$

and

$$p_h(x,\{y:|y-x|\geq \xi\}) = \int_{|z|\geq \frac{\xi}{\epsilon}} \alpha(x,z,\epsilon)q(x,z)dz.$$

For condition (4.3)

$$\begin{aligned}\sup_{|x|\leq r}\frac{1}{h}p_h(x,\{y:|y-x|\geq\xi\}) = \sup_{|x|\leq r}\frac{1}{\epsilon^2}\int_{|z|>\frac{\xi}{\epsilon}}\alpha(x,z,\epsilon)q(x,z)dz &\leq \sup_{|x|\leq r}\frac{1}{\epsilon^2}\int_{|z|\geq\frac{\xi}{\epsilon}} q(x,z)dz \\ &= \sup_{|x|\leq r}\frac{1}{\epsilon^2}\mathsf{erfc}\left(\frac{\sqrt{D(x)}\xi}{\epsilon}\right)\to 0.\end{aligned}$$

Since we want to show $b_h^2(x)$ and $a_h(x)$ are converging to some continuous functions, we are going to show the equicontinuity and then apply the Arzelà-Ascoli theorem. Recall that a family of continuous functions G on compact metric space (S,λ) is equicontinuous if and only if

$$\forall\epsilon>0,\forall x\in S,\exists\delta>0,\forall y\in S,\forall g\in G,\lambda(x,y)<\delta\Rightarrow|g(x)-g(y)|<\epsilon.$$

For condition (4.2), firstly we show the equicontinuity of $b_h^2(x)$:
For $\forall\,|x|\leq r$, $\forall\eta>0$, let $B(x,\gamma)$ be an open ball centered at x with radius $\gamma>0$ such that for $\forall y\in B(x,\gamma)$, we have

$$\begin{aligned}\left|b_h^2(x)-b_h^2(y)\right| &= \left|\int_{|z|\leq\frac{r}{\epsilon}} zz^T(\alpha(x,z,\epsilon)q(x,z)-\alpha(y,z,\epsilon)q(y,z))dz\right| \\ \leq \left|\int_{|z|\leq\frac{r}{\epsilon}} zz^T(\alpha(x,z,\epsilon)-\alpha(y,z,\epsilon))q(x,z)dz\right| &+ \left|\int_{|z|\leq\frac{r}{\epsilon}} zz^T\alpha(y,z,\epsilon)(q(x,z)-q(y,z))dz\right|\end{aligned}$$

Then, using Lemma 6, the first term on the right hand side of the inequality has the following upper bound.

$$\left|\int_{|z|\leq\frac{r}{\epsilon}} zz^T(\alpha(x,z,\epsilon)-\alpha(y,z,\epsilon))q(x,z)dz\right|$$
$$\leq \int_{|z|\leq\frac{r}{\epsilon}} |z|^2\left((|g_1(z)|+|\nabla D(x)|\,|g_2(z)|+|g_3(z)|\,\epsilon)\epsilon\,|y-x|\right)q(x,z)dz \quad\leq\quad K_1(x)\,|y-x|$$

where $K_1(x)$ is a continuous function of x independent of h or ϵ. Using Lemma 5, the second term has upper bound,

$$\left|\int_{|z|\leq\frac{r}{\epsilon}} zz^T\alpha(y,z,\epsilon)(q(x,z)-q(y,z))dz\right| \leq \int_{|z|\leq\frac{r}{\epsilon}} |z|^2\,|q(x,z)-q(y,z)|\,dz$$
$$\leq M\int_{|z|\leq\frac{r}{\epsilon}} |z|^2\,e^{-\frac{z^2}{4K}}\left(\left|\frac{\nabla D(x)}{D^2(x)}\right|+(1+|\nabla D(x)|)\,|y-x|\right)|x-y|\,dz \leq K_2(x)\,|y-x|$$

where $K_2(x)$ is also a continuous function of x independent of h or ϵ. As a result,

$$\left|b_h^2(x)-b_h^2(y)\right| \leq (K_1(x)+K_2(x))\,|y-x| \leq (K_1(x)+K_2(x))\gamma$$

So we can choose γ small enough to make the difference of $b_h^2(x)-b_h^2(y)$ smaller than η for all h. Hence, $b_h^2(x)$ is equicontinuous for all $|x|<r$. For a fixed x, since $0\leq\alpha\leq 1$, the integrand in $b_h^2(x)$ is dominated

$$\left|zz^T(\alpha(x,z,\epsilon))q(x,z)\right| \leq \left|zz^Tq(x,z)\right|$$

Since

$$b^2(x)=\int_{\mathbb{R}^d} zz^Tq(x,z)dz = 2D(x)I_d$$

Therefore, by the Arzelà-Ascoli Theorem [3, p. 190], $b_h^2(x)\to 2D(x)I_d$ uniformly in x as $h\to 0$. As a result

$$\sup_{|x|\leq r}\left|b_h^2(x)-b^2(x)\right|\to 0$$

For condition (4.1), similarly we show the equicontinuity of $a_h(x)$:
For $\forall\,|x|\leq r$, $\forall\eta>0$, let $B(x,\gamma)$ be an open ball centered at x with radius $\gamma>0$ such that for $\forall y\in B(x,\gamma)$, we have

$$|a_h(x)-a_h(y)| = \frac{1}{\epsilon}\left|\int_{|z|\leq\frac{r}{\epsilon}} z(\alpha(x,z,\epsilon)q(x,z)-\alpha(y,z,\epsilon)q(y,z))dz\right|$$
$$\leq \frac{1}{\epsilon}\left|\int_{|z|\leq\frac{r}{\epsilon}} z(\alpha(x,z,\epsilon)-\alpha(y,z,\epsilon))q(x,z)dz\right| + \frac{1}{\epsilon}\left|\int_{|z|\leq\frac{r}{\epsilon}} z\alpha(y,z,\epsilon)(q(x,z)-q(y,z))dz\right|$$

Then, using Lemma 6, the first term in the inequality has the upper bound

$$\frac{1}{\epsilon}\left|\int_{|z|\le\frac{r}{\epsilon}} z(\alpha(x,z,\epsilon)-\alpha(y,z,\epsilon))q(x,z)dz\right| \le$$
$$\int_{|z|\le\frac{r}{\epsilon}} |z|\left(\left(|g_1(z)|+|\nabla D(x)|\,|g_2(z)|+|g_3(z)|\,\epsilon\right)|y-x|\right) q(x,z)dz$$
$$\le K_3(x)\,|y-x|$$

where $K_3(x)$ is a continuous function of x independent of h or ϵ. Using Lemma 5 and Lemma 7, the second term has the upper bound

$$\frac{1}{\epsilon}\left|\int_{|z|\le\frac{r}{\epsilon}} z\alpha(y,z,\epsilon)(q(x,z)-q(y,z))dz\right| = \frac{1}{\epsilon}\left|\int_{|z|\le\frac{r}{\epsilon}} z(\alpha(y,z,\epsilon)-1)(q(x,z)-q(y,z))dz\right|$$
$$\le \int_{|z|\le\frac{r}{\epsilon}} |z|\,g(y,z)\,|q(x,z)-q(y,z)|\,dz \le K_4(x)\,|y-x|$$

where $K_4(x)$ is also a continuous function of x independent of h or ϵ. Even though we have $g(y,z)$ not depending on x in the above integral, since $g(y,z)$ is a polynomial while y is close to x. Therefore it is always possible to find a larger function $K_4(x)$ depending on x as the upper bound as long as x is in some compact set.
As a result,

$$|a_h(x)-a_h(y)| \le (K_3(x)+K_4(x))\gamma$$

Hence, $a_h(x)$ is equicontinuous for all $|x|<r$. On the other hand, using Lemma 7, $a_h(x)$ can be shown uniformly bounded by some constant U for $|x|<r$

$$|a_h(x)| = \frac{1}{\epsilon}\left|\int_{|z|\le\frac{r}{\epsilon}} z(\alpha(x,z,\epsilon)-1)q(x,z)dz\right| \le \left|\int_{|z|\le\frac{r}{\epsilon}} zg(x,z)q(x,z)dz\right| \le U$$

Therefore, by the Arzela-Ascoli Theorem [3, p. 190], $a_h(x)$ is relatively compact. For any convergent subsequence $\{a_{h_j}(x)\}$, $h_j\to 0$, it is converging uniformly to some continuous function $\hat{a}(x)$. Therefore, after applying Theorem 6. This subsequence is converging to a solution to a martingale problem with coefficients $\hat{a}, b^2$, i.e., the solution from the numerical method is converging to an Itô diffusion (A.2) with drift $\hat{a}$ and diffusion b. To show furthermore that $\hat{a}=a$ is unique for all convergent subsequences, we shall use the detailed balance condition. As shown in Section 3.3, the Metropolized integrator preserves the exact equilibrium density ρ_{eq} and satisfies the detailed balance condition with respect to ρ_{eq}. Let $S_{xy}=S_x\times S_y$ where S_x and S_y are measurable sets in $\mathbb{R}^d$. Then the detailed balance condition for the transitional probability distribution P_h^t of the numerical method is

$$\int_{S_{xy}} \rho_{\mathsf{eq}}(x)P_h^t(x,dy)dx = \int_{S_{yx}} \rho_{\mathsf{eq}}(x)P_h^t(x,dy)dx$$

which means the transitions from S_x to S_y is the same as the transitions from S_y to S_x, for $t = nh$, $n = 1..N$. Here $P_h^t(x, dy)$ denotes the transitional probability density from state x to dy after a fixed amount of time $t = nh$ for time step length h. By a change of variable, we obtain

$$\int_{S_{xy}} \rho_{\mathsf{eq}}(x) P_h^t(x, dy)dx = \int_{S_{xy}} \rho_{\mathsf{eq}}(y) P_h^t(y, dx)dy \tag{4.4}$$

The weak convergence of the numerical solution to diffusions implies that the distribution $\rho_{\mathsf{eq}}(x)P_h^t(x, dy)dx \Rightarrow \rho_{\mathsf{eq}}(x)P^t(x, dy)dx$ weakly. By the equivalent definitions of weak convergence [37, p. 117] (see Appendix **??**), for any measurable set Σ such that the boundary $\partial\Sigma$ has zero measure under $\rho_{\mathsf{eq}}(x)P^t(x, dy)dx$, we have

$$\lim_{h\to 0} \int_\Sigma \rho_{\mathsf{eq}}(x) P_h^t(x, dy)dx = \int_\Sigma \rho_{\mathsf{eq}}(x) P^t(x, dy)dx.$$

Since the diffusion coefficient is bounded away from zero, the density of P^t satisfies a second order parabolic equation(the Fokker-Planck equation). As a result, $P^t(x, dy) = p^t(x, y)dy$ is Lebesgue measurable and $p^t(x, y)$ is continuous both in x, y for all $t > 0$. Therefore, by taking $h \to 0$ and letting S_{xy} have boundaries with Lebesgue measure zero, (4.4) becomes

$$\int_{S_{xy}} \rho_{\mathsf{eq}}(x) P^t(x, dy)dx = \int_{S_{xy}} \rho_{\mathsf{eq}}(y) P^t(y, dx)dy = \int_{S_{yx}} \rho_{\mathsf{eq}}(x) P^t(x, dy)dx. \tag{4.5}$$

Since diffusions have continuous paths, (4.5) suffices to show that the limit process have no net probability flux under the density ρ_{eq} [14, p. 119], i.e. $J = 0$ in (2.4). From the discussion in 2.3, $J = 0$ implies that the drift has the unique expression $\hat{a}(x) = \nabla D(x) + D(x)\nabla \ln \rho_{\mathsf{eq}}(x) = a(x)$. Thus, all limits of the convergent subsequence are the same as $a(x)$. Therefore, as $h \to 0$, there is a continuous function $a(x)$,

$$\sup_{|x|\le r} |a_h(x) - a(x)| \to 0$$

and the limiting diffusion solves (2.6). □

Lemma 4. *With the assumptions in Theorem 7, we have the following estimates for the smoothness of D, $\frac{\nabla D}{D}$ and $\frac{\nabla D}{D^2}$*

1. $$|D(x) - D(y)| \le L_1 \frac{\sqrt{2D(x)} + \sqrt{2D(y)}}{2} |x - y|$$

2. $$\left| \frac{\nabla D(x)}{D(x)} - \frac{\nabla D(y)}{D(y)} \right| \le (C_1 + C_2 |\nabla D(x)|) |y - x|$$

3.

$$\left|\frac{\nabla D(x)}{D^2(x)} - \frac{\nabla D(y)}{D^2(y)}\right| \leq (C_3 + C_4 |\nabla D(x)|) |y - x|$$

Proof. For the first inequality, we have

$$|D(x) - D(y)| = \left|\frac{\sqrt{2D(x)} + \sqrt{2D(y)}}{2}(\sqrt{2D(x)} - \sqrt{2D(y)})\right| \leq L_1 \frac{\sqrt{2D(x)} + \sqrt{2D(y)}}{2} |x - y|$$

For the second inequality, by direct calculation,

$$\begin{aligned}
\left|\frac{\nabla D(x)}{D(x)} - \frac{\nabla D(y)}{D(y)}\right| &\leq \frac{1}{D(x)D(y)} |D(y)\nabla D(x) - D(x)\nabla D(y)| \\
&\leq \frac{1}{D(x)D(y)} (|D(y)\nabla D(x) - D(x)\nabla D(x)|) + \frac{1}{D(x)D(y)} (|D(x)\nabla D(x) - D(x)\nabla D(y)|) \\
&\leq \frac{1}{D(x)D(y)} \left(\left|\nabla D(x) L_1 \frac{\sqrt{2D(x)} + \sqrt{2D(y)}}{2} |x - y|\right|\right) + \frac{1}{D(x)D(y)} (|D(x)L_3(x - y)|) \\
\leq \frac{1}{\sqrt{2D(x)}D(y)} (|\nabla D(x) L_1 |x - y||) + \frac{1}{D(x)\sqrt{2D(y)}} (|\nabla D(x) L_1 |x - y||) + \frac{1}{D(y)} (|L_3(x - y)|) \\
&\leq (C_1 + C_2 |\nabla D(x)|) |y - x|
\end{aligned}$$

Therefore $\frac{\nabla D(x)}{D(x)}$ is locally Lipschitz. For the third inequality, by direct calculation,

$$\begin{aligned}
&\left|\frac{\nabla D(x)}{D^2(x)} - \frac{\nabla D(y)}{D^2(y)}\right| \leq \frac{1}{D^2(x)D^2(y)} \left|D^2(y)\nabla D(x) - D^2(x)\nabla D(y)\right| \\
&\leq \frac{1}{D^2(x)D^2(y)} \left(\left|D^2(y)\nabla D(x) - D^2(x)\nabla D(x)\right|\right) + \frac{1}{D^2(x)D^2(y)} \left(\left|D^2(x)\nabla D(x) - D^2(x)\nabla D(y)\right|\right) \\
&\leq \frac{1}{D^2(x)D^2(y)} \left(\left|\nabla D(x) L_1 \frac{(D(x) + D(y))(\sqrt{2D(x)} + \sqrt{2D(y)})}{2} |x - y|\right|\right) \\
&+ \frac{1}{D^2(x)D^2(y)} \left(\left|D^2(x)L_3(x - y)\right|\right) \\
&\leq (C_3 + C_4 |\nabla D(x)|) |y - x|
\end{aligned}$$

Therefore $\frac{\nabla D(x)}{D^2(x)}$ is locally Lipschitz. □

Lemma 5. *With the assumptions in Theorem 7,* $|q(x, z) - q(y, z)|$ *has the following upper bound,*

$$|q(x, z) - q(y, z)| \leq M e^{-\frac{z^2}{4K}} \left(\left|\frac{\nabla D(x)}{D^2(x)}\right| + (1 + |\nabla D(x)|) |y - x|\right) |x - y|$$

for some positive constant M*.*

Proof. Using mean value theorem and by direct calculation

$$q(x, z) - q(y, z) = \left(-\frac{d}{2}(4\pi D(\xi))^{-\frac{d}{2}} e^{-\frac{z^2}{4D(\xi)}} \frac{\nabla D(\xi)}{D(\xi)} + (4\pi D(\xi))^{-\frac{d}{2}} e^{-\frac{z^2}{4D(\xi)}} \frac{z^2}{4} \frac{\nabla D(\xi)}{D^2(\xi)}\right) \cdot (x - y)$$

Since we have $\inf D = K$ and $\frac{\nabla D}{D}$, $\frac{\nabla D}{D^2}$ satisfying Lipschitz conditions as shown in Lemma 4,

$$\begin{aligned}|q(x,z)-q(y,z)| &\le \left|(\frac{d}{2}(4\pi K)^{-\frac{d}{2}}e^{-\frac{z^2}{4K}}\left(\left|\frac{\nabla D(x)}{D(x)}\right|+(C_1+C_2\left|\nabla D(x)\right|)\left|y-x\right|\right)\right|\cdot|x-y| \\ &+\left|(4\pi K)^{-\frac{d}{2}}e^{-\frac{z^2}{4K}}\frac{z^2}{4}\left(\left|\frac{\nabla D(x)}{D^2(x)}\right|+(C_3+C_4\left|\nabla D(x)\right|)\left|y-x\right|\right)\right|\cdot|x-y| \\ &\le Me^{-\frac{z^2}{4K}}\left(\left|\frac{\nabla D(x)}{D^2(x)}\right|+(1+\left|\nabla D(x)\right|)\left|y-x\right|\right)|x-y|\end{aligned}$$

For some constant $M>0$ □

Lemma 6. *With the assumptions in Theorem 7, $\alpha(x,z,\epsilon)$ satisfies a Lipschitz condition in x*

$$|\alpha(x,z,\epsilon)-\alpha(y,z,\epsilon)| \le (|g_1(z)|+|\nabla D(x)|\,|g_2(z)|+|g_3(z)|\,\epsilon)\epsilon\,|y-x|$$

with g_1,g_2,g_3 are polynomials.

Proof. Let

$$\alpha(x,z,\epsilon)=\min\left(\frac{q(x+\epsilon z)\rho_{\mathsf{eq}}(x+\epsilon z)}{q(x,z)\rho_{\mathsf{eq}}(x)},1\right)=\min(\hat{\alpha}(x,z,\epsilon),1)$$

then

$$\begin{aligned}|\alpha(x,z,\epsilon)-\alpha(y,z,\epsilon)| &= \left|\min(e^{\ln\hat{\alpha}(x,z,\epsilon)},1)-\min(e^{\ln\hat{\alpha}(y,z,\epsilon)},1)\right| \\ &\le |\ln\hat{\alpha}(x,z,\epsilon)-\ln\hat{\alpha}(y,z,\epsilon)|\end{aligned}$$

Since,

$$\begin{aligned}\ln\hat{\alpha}(x,z,\epsilon) &= \ln q(x+\epsilon z,z)-\ln q(x,z)+\ln\rho_{\mathsf{eq}}(x+\epsilon z)-\ln\rho_{\mathsf{eq}}(x) \\ &= \int_0^\epsilon \frac{z\cdot\nabla q(x+\zeta z,z)}{q(x+\zeta z,z)}d\zeta+\int_0^\epsilon\frac{z\cdot\nabla\rho_{\mathsf{eq}}(x+\zeta z)}{\rho_{\mathsf{eq}}(x+\zeta z)}d\zeta \\ &= \int_0^\epsilon -\frac{d}{2}\cdot\frac{z\cdot\nabla D(x+\zeta z)}{D(x+\zeta z)}-\frac{z^2(z\cdot\nabla D(x+\zeta z))}{4D^2(x+\zeta z)}d\zeta+\int_0^\epsilon\frac{z\cdot\nabla\rho_{\mathsf{eq}}(x+\zeta z)}{\rho_{\mathsf{eq}}(x+\zeta z)}d\zeta\end{aligned}$$

$$\begin{aligned}|\ln\hat{\alpha}(x,z,\epsilon)-\ln\hat{\alpha}(y,z,\epsilon)| &\le \left|\int_0^\epsilon -\frac{d}{2}\cdot z\cdot\left(\frac{\nabla D(x+\zeta z)}{D(x+\zeta z)}-\frac{\nabla D(y+\zeta z)}{D(y+\zeta z)}\right)d\zeta\right| \\ &+\left|\int_0^\epsilon -z^2\left(z\cdot\left(\frac{\nabla D(x+\zeta z)}{4D^2(x+\zeta z)}-\frac{\nabla D(y+\zeta z))}{4D^2(y+\zeta z)}\right)\right)d\zeta\right| \\ &+\left|\int_0^\epsilon z\cdot\left(\frac{\nabla\rho_{\mathsf{eq}}(x+\zeta z)}{\rho_{\mathsf{eq}}(x+\zeta z)}-\frac{\nabla\rho_{\mathsf{eq}}(y+\zeta z)}{\rho_{\mathsf{eq}}(y+\zeta z)}\right)d\zeta\right|\end{aligned}$$

Therefore, using Lemma 4

$$\left|\frac{\nabla D(x+\zeta z)}{D(x+\zeta z)}-\frac{\nabla D(y+\zeta z))}{D(y+\zeta z)}\right| \le (C_1+C_2\left|\nabla D(x+\zeta z)\right|)\left|y-x\right| \le (C_1+C_2\left|\nabla D(x)\right|+L_3\left|\zeta z\right|))\left|y-x\right|$$

$$\left|\frac{\nabla D(x+\zeta z)}{4D^2(x+\zeta z)} - \frac{\nabla D(y+\zeta z))}{4D^2(y+\zeta z)}\right| \leq (C_3 + C_4 \left|\nabla D(x+\zeta z)\right|) \left|y-x\right| \leq (C_3 + C_4 \left|\nabla D(x)\right| + L_3 \left|\zeta z\right|)) \left|y-x\right|$$

$$\left|\frac{\nabla \rho_{\mathsf{eq}}(x+\zeta z)}{\rho_{\mathsf{eq}}(x+\zeta z)} - \frac{\nabla \rho_{\mathsf{eq}}(y+\zeta z)}{\rho_{\mathsf{eq}}(y+\zeta z)}\right| \leq L_2 \left|y-x\right|$$

Hence,

$$|\alpha(x,z,\epsilon) - \alpha(y,z,\epsilon)| \leq (|g_1(z)| + |\nabla D(x)|\,|g_2(z)| + |g_3(z)|\,\epsilon)\epsilon\,|y-x|$$

□

Lemma 7. *With the assumptions in Theorem 7,* $(\alpha(x,z,\epsilon)-1)/\epsilon$ *is uniformly bounded by a polynomial in* x, z *for all* ϵ

$$|\alpha(x,z,\epsilon) - 1| \leq g(x,z)\epsilon$$

with g *a polynomial in* x, z*.*

Proof. The proof of the uniform boundness is similar to the proof of the Lipschitz condition in Lemma 6 Let

$$\alpha(x,z,\epsilon) = \min\left(\frac{q(x+\epsilon z)\rho_{\mathsf{eq}}(x+\epsilon z)}{q(x,z)\rho_{\mathsf{eq}}(x)}, 1\right) = \min(\hat{\alpha}(x,z,\epsilon), 1)$$

then

$$\begin{aligned} |\alpha(x,z,\epsilon) - 1| &= \left|\min(e^{\ln\hat{\alpha}(x,z,\epsilon)}, 1) - \min(e^0, 1)\right| \\ &\leq |\ln\hat{\alpha}(x,z,\epsilon) - 0| \end{aligned}$$

Since,

$$\begin{aligned} \ln\hat{\alpha}(x,z,\epsilon) &= \ln q(x+\epsilon z, z) - \ln q(x,z) + \ln\rho_{\mathsf{eq}}(x+\epsilon z) - \ln\rho_{\mathsf{eq}}(x) \\ &= \int_0^\epsilon \frac{z\cdot\nabla q(x+\zeta z, z)}{q(x+\zeta z, z)} d\zeta + \int_0^\epsilon \frac{z\cdot\nabla\rho_{\mathsf{eq}}(x+\zeta z)}{\rho_{\mathsf{eq}}(x+\zeta z)} d\zeta \\ = \int_0^\epsilon -\frac{d}{2}\cdot\frac{z\cdot\nabla D(x+\zeta z)}{D(x+\zeta z)} - \frac{z^2(z\cdot\nabla D(x+\zeta z))}{4D^2(x+\zeta z)} d\zeta &+ \int_0^\epsilon \frac{z\cdot\nabla\rho_{\mathsf{eq}}(x+\zeta z)}{\rho_{\mathsf{eq}}(x+\zeta z)} d\zeta \end{aligned}$$

Therefore, using Lemma 4

$$\left|\frac{\nabla D(x+\zeta z)}{D(x+\zeta z)} - \frac{\nabla D(x))}{D(x)}\right| \leq (C_1 + C_2\left|\nabla D(x)\right|)\left|z\right|\zeta \leq (C_1 + C_2\left|\nabla D(x)\right|)\left|z\right|\epsilon$$

$$\left|\frac{\nabla D(x+\zeta z)}{4D^2(x+\zeta z)} - \frac{\nabla D(x))}{4D^2(x)}\right| \leq (C_3 + C_4\left|\nabla D(x)\right|)\left|z\right|\zeta \leq (C_3 + C_4\left|\nabla D(x)\right|)\left|z\right|\epsilon$$

$$\left| \frac{\nabla \rho_{\mathsf{eq}}(x+\zeta z)}{\rho_{\mathsf{eq}}(x+\zeta z)} - \frac{\nabla \rho_{\mathsf{eq}}(x)}{\rho_{\mathsf{eq}}(x)} \right| \leq L_2 \left| z) \right| \zeta \leq L_2 \left| z) \right| \epsilon$$

Hence,

$$|\alpha(x,z,\epsilon) - 1| \leq g(x,z)\epsilon$$

□

Chapter 5

Summary and Conclusion

This thesis builds upon the work of my graduate research in Simon Fraser University with my senior supervisor Dr. Paul Tupper. Our main goals are to build up 1) a new formulation of the SDE to correctly model diffusive systems with variable diffusion coefficients in the absence of external force, 2) a numerical method that utilizes the new formulation to solve the SDE.

We firstly show that for diffusive systems with variable diffusion coefficients, the physical systems cannot be fully described solely by the diffusion coefficients even if there are no external force. The justifications are from both theoretical discussions of different interpretations of stochastic integrals and a concrete microscopic computational simulation of the Lorentz gas model. Therefore, a drift term which is not from external force should come with the diffusion. The computational simulation also reveals that there is a family of parameters producing the same diffusion coefficients but giving different long time behaviors. Hence, we suggest to choose the equilibrium distribution combined with the detailed balance condition as well as the diffusion coefficient to describe a diffusive system. The underlying drift is uniquely determined by the diffusion coefficient and the equilibrium density with the detailed balance condition.

Then we propose a numerical method that takes advantage of the new formulation in terms of the diffusion coefficient and the equilibrium density. The numerical method uses the Euler-Maruyama method for the drift-free SDE to give a proposal step and then uses Metropolis-Hastings algorithm to introduce rejections which generate drifts implicitly. The detailed balance condition is automatically satisfied by the Metropolis-Hastings algorithm. Therefore, no explicit calculation of the drift term is required. As a result, we are able to deal with the cases where the drift may have singularity when the diffusion coefficient and the equilibrium density are not smooth enough. Another important advantage is that the equilibrium density is preserved exactly by the numerical method which means that the numerical method is capable of simulating long time behavior. We show that the numerical method is $\frac{1}{2}$ order accurate by error analysis for smooth coefficients and it is also validated to be convergent by computational experiments even for coefficients that are less smooth

than those stated in the conditions of the convergence theorem and for equilibrium densities that are not normalizable.

Finally we use diffusion approximations as an alternative way to justify the weak convergence of the numerical method. This way of proof, as a complement to the previous error analysis, shows the weak convergence without giving any rate and requires less for smoothness of the coefficients. Since the well-posedness of the corresponding martingale problems can be obtained under the conditions of diffusion coefficients being continuous and positive and drift coefficients being measurable, we can use diffusion approximations and the martingale problem to investigate the weak convergence of the numerical method for discontinuous coefficients in the future study.

Bibliography

[1] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On logarithmic Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Partial Differential Equations*, 26((1-2)):43–100, 2001.

[2] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., 2nd edition edition, 1999.

[3] Adam Bobrowski. *Functional Analysis for Probability and Stochastic Processes: An Introduction*. Cambridge University Press, 2005.

[4] Nawaf Bou-Rabee, Aleksandar Donev, and Eric Vanden-Eijnden. Metropolis integration schemes for self-adjoint diffusions. 2014.

[5] Nawaf Bou-Rabee and Eric Vanden-Eijnden. Pathwise accuracy and ergodicity of Metropolized integrator for sdes. *Communications on Pure and Applied Mathematics*, 63:655–696, 2010.

[6] C. Bruin. Logarithmic terms in the diffusion coefficient for the lorentz gas. *Phys. Rev. Lett.*, 29:1670–1674, 1972.

[7] L. A. Bunimovich and Y. G. Sinaĭ. Statistical properties of Lorentz gas with periodic configuration of scatterers. *Comm. Math. Phys.*, 78(4):479–497, 1980/81.

[8] Siddhartha Chib and Edward Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[9] C. P. Dettmann and E. G. D. Cohen. Microscopic chaos and diffusion. *J. Statist. Phys.*, 101(3-4):775–817, 2000.

[10] C.P. Dettmann. The lorentz gas: a paradigm for nonequilibrium stationary states. In D. Szász, editor, *Hard Ball Systems and the Lorentz Gas*, volume 101 of *Encyclopaedia of Mathematical Sciences*, pages 315–365. Springer Berlin Heidelberg, 2000.

[11] M. Ernst and A. Weyland. Long time behaviour of the velocity auto-correlation function in a Lorentz gas. *Phys. Lett. A*, 34(1):39–40, 1971.

[12] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley, 1986.

[13] Pierre Étoré. On random walk simulation of one-dimensional diffusion processes with discontinuous coefficients. *Electron. J. Probab.*, 11:no. 9, 249–275 (electronic), 2006.

[14] C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natrual sciences*. Berlin, Germany: Springer, 3rd edn edition, 2004.

[15] I. I. Gihman and A. V. Skorohod. *Stochastic Differential Equations*. 1970.

[16] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series and Products*. Elsevier/Academic Press, Amsterdam, 2007.

[17] D. Hall and M. Hoshino. Effects of marcromolecular crowding on intracellular diffusion from a single particle persperctive. *Biophys. Rev.*, 2:39–53, 2010.

[18] Desmond J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

[19] Werner Horsthemke and R. Lefever. *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology*. Springer-Verlag Berlin Heidelberg, 2006.

[20] Soren Fiig Jarner and Ernst Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and their Applications*, 85(1):341–361, 2000.

[21] N. Korabel and E. Barkai. Boundary conditions of normal and anomalous diffusions from thermal equilibrium. *Phys. Rev. E*, 83:051113, 2011.

[22] Eric M. LaBolle, Jeremy Quastel, Graham E. Fogg, and Janko Gravner. Diffusion processes in composite porous media and their numerical integration by random walks: Generalized stochastic differential equations with discontinuous coefficients. *Water Resources Research*, 36(3):651–662, 2000.

[23] P. Lançon, G. Batrouni, L. Lobry, and N. Ostrowsky. Drift without flux: Brownian walker with a space-dependent diffusion coefficient. *Europhysics Letters*, 54:28–34, 2001.

[24] A. W. C. Lau and T. C. Lubensky. State-dependent diffusion: Thermodynamic consistency and its path integral formulation. *Phys. Rev. E*, 76(1):011123, 2007.

[25] Antoine Lejay and Géraldine Pichot. Simulating diffusion processes in discontinuous media: A numerical scheme with constant time steps. *Journal of Computational Physics*, 231(21):7299 – 7314, 2012.

[26] H. A. Lorentz. The motion of electrons in metallic bodies. *Proc. R. Acad. Amst.*, 7:438–453, 1905.

[27] P. Markowich and C. Villani. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Mat. Contemp.*, 19:1–29, 2000.

[28] Miguel Martinez and Denis Talay. One-dimensional parabolic diffraction equations: pointwise estimates and discretization of related stochastic differential equations with weighted local times. *Electron. J. Probab.*, 17:no. 27, 30, 2012.

[29] Grigori Noah Milstein and Michael V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer, 2004.

[30] Bernt Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer-Verlag.

[31] H. G. Othmer and A. Stevens. Aggregation, blowup, and collapse: the ABC's of taxis in reinforced random walks. *SIAM J. Appl. Math.*, 57:1044–1081, 1997.

[32] O. Ovaskainen and S. J. Cornell. Biased movement at a boundary and conditional occupancy times for diffusion processes. *J. Appl. Probab.*, 40(3):557–580, 2003.

[33] Frederick Reif. *Fundamentals of statistical and thermal physics*. New York, NY: McGraw-Hill, 1965.

[34] D. Ridgway, G. Broderick, A. Lopez-Campistrous, M. Ru'aini, P. Winter, M. Hamilton, P. Boulanger, A. Kovalenko, and M. J. Ellison. Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. *Biophys. J.*, 94:3748–3759, 2008.

[35] R.Klages and C.Dellago. Density-dependent diffusion in the periodic Lorentz gas. *J. Stat. Phys.*, 101:145–159, 2000.

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

[37] Jeffrey S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific Publishing Company, 2006.

[38] D. Sanders. *Deterministic Diffusion in Periodic Billiard Models*. PhD thesis, University of Warwick, 2005.

[39] Eric D. Siggia, Jennifer Lippincott-Schwartz, and Stefan Bekiranov. Diffusion in inhomogeneous media: Theory and simulations applied to whole cell photobleach recovery. *Biophysical Journal*, 79(4):1761 – 1770, 2000.

[40] J. G. Sinaĭ. Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Uspehi Mat. Nauk*, 25(2):141–192, 1970.

[41] A. F. M. Smith and G. O. Roberts. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(1):3–23, 1993.

[42] D.W. Stroock and S.R.S Varadhan. *Multidimensional Diffusion Process*. Springer-Verlag, 1979.

[43] Paul. F. Tupper and Xin Yang. A paradox of state-dependent diffusion and how to resolve it. *Proceedings of the royal society A*, 468:3864–3881, 2012.

[44] T. E. Turner and S. Schnell annd K. Burrage. Stochastic approaches for modelling in vivo reactions. *Comp. Biol. Chem.*, 28:165–178, 2004.

[45] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library. North-Holland, 3rd edition, 2007.

[46] J. van Leeuwen and A. Weijland. Non-analytic density behaviour of the diffusion coefficient of a Lorentz gas i. divergencies in the expansion in powers in the density. *Physica*, 36(3):457–490, 1967.

[47] G. Volpe, L. Helden, T. Brettschneider, J. Wehr, and C. Bechinger. Influence of noise on force measurements. *Phys. Rev. Lett.*, 104(17):170602, 2010.

Appendix A

Preliminaries

Here are some basic materials in stochastic processes and SDEs. Most contents can be found in [37, 30].

A.1 Measure-theoretic Probability

Probability is some number (between 0 and 1) denoting the possibility that some event happens. To represent the idea mathematically, Measure-theoretic Probability assumes that the events are the subsets of some non-empty set Σ. Then the probability of one event $A \in \Sigma$ happening is $\mathbb{P}(A)$ where $\mathbb{P}$ is a map from subsets of Σ to a real number $p \in [0, 1]$. Therefore, the idea of probability naturally grows upon the measure theory.

Definition 3. *A probability space or a probability triple is* $(\Omega, \mathcal{F}, \mathbb{P})$ *with following property.*

- Ω *is a non-empty set called the sample space;*
- $\mathcal{F}$ *is the* σ*-algebra which consists of subsets of* Ω*, containing* Ω *itself and the empty set* $\emptyset$*;*
- $\mathbb{P}$ *is a mapping from* $\mathcal{F}$ *to* $[0, 1]$*, with* $\mathbb{P}(\emptyset) = 0$ *and* $\mathbb{P}(\Omega) = 1$*, such that* $\mathbb{P}$ *is countably additive.*

This definition allows us to study the possibility of any event A as long as A is measurable ($A \in \mathcal{F}$). A more interesting to study is one experiment that is built on the event $A \in \mathcal{F}$. By thinking of the sample space Ω as the set of all possible outcomes of some experiment, then a random variable assigns a numerical value to each of these outcomes.

Definition 4. *Given a probability triple* $(\Omega, \mathcal{F}, \mathbb{P})$*, a random variable is a function* X *from* Ω *to the real numbers* $\mathbb{R}$*, such that*

$$\{\omega \in \Omega;\ X(\omega) \leq x\} \in \mathcal{F}, \quad x \in \mathbb{R}$$

Therefore, a random variable $X(\omega)$ is a measurable function on $(\Omega, \mathcal{F}, \mathbb{P})$. This definition could be easily extended to X taking values in multiple dimensions by thinking X as a measurable function from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}^d, \mathcal{B}, \lambda)$.

A.2 Stochastic Processes

To study the relation or evolution of random variables, we introduce a collection of random variables called stochastic process.

Definition 5. *A stochastic process is a parametrized collection of random variables*

$$\{X_t\}_{t \in T}$$

defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *and assuming values in* $\mathbb{R}^d$

Here the parameter T is a subset in $\mathbb{R}$ and is usually describing the time. T could be discrete ($T = \{i; i \in \mathbb{N}\}$) or continuous ($T = [0, \infty)$). Note that for each $t \in T$ fixed we have a random variable

$$\omega \to X_t(\omega); \;\; \omega \in \Omega.$$

One the other hand, fixing $\omega \in \Omega$ we can consider the function

$$t \to X_t(\omega); \;\; t \in T.$$

We shall call this a path or trajectory of X_t. Let's look at a famous example of stochastic process: Brownian motion or Wiener Process.

Definition 6. *Brownian motion (Wiener Process) is a continuous time stochastic process* $B_t, \; t \in [0, \infty)$ *that satisfies,*

- B_t *is a Gaussian process: for all* $0 \leq t_1 \leq t_2 \cdots \leq t_k$ *the random variable* $Z = (B_{t_1}, \cdots, B_{t_k})\mathbb{R}^{nk}$ *has a multinormal distribution;*

- *independent increments: for all* $0 \leq t_1 \leq t_2 \cdots \leq t_k$, $B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_k} - B_{t_{k-1}}$ *are independent;*

- *continuous paths:* $\mathbb{P}(\{\omega; B_t(\omega)$ *is continuous*$\}) = 1$

The construction of Brownian motion is not trivial. [30, p. 11] provides one way to show the existence and continuous paths using Kolmogorov's extension theorem and Kolmogorov's continuity theorem.

A.3 Martingales

Martingales are important stochastic processes which are originally used to describe fair games. The 'fair' is shown through conditional expectations which says that expectation of a stochastic process X_t at time t conditioned on the process at some previous time $t' < t$, (use history information at and before t' to predict the future value at t) is $X_{t'}$: $\mathbb{E}(X_t|X_{\tilde{t}}, (\tilde{t} \leq t')) = X_{t'}$. Since the conditional expectation conditioned on random variables is defined by the conditional expectation conditioned on the σ-algebra generated by the random variables. The information is equivalently described by an increasing family of σ-algebras which is called a filtration.

Definition 7. *A filtration on some measurable space $(\Omega, \mathcal{F})$ is a familiy $\mathcal{M} = \{\mathcal{M}_t\}(t \geq 0)$ of σ-algebras $\mathcal{M}_t \subset \mathcal{F}$ such that*

$$0 \leq s < t \Rightarrow \mathcal{M}_s \subset \mathcal{M}_t$$

And a process X_t is said to be adapted to a filtration $\mathcal{M}$ if and only if X_s is $\mathcal{M}_s$ measurable for all s.

Definition 8. *An n-dimensional stochastic process $M_t, (t \geq 0)$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a martingale with respect to a filtration $\mathcal{M}_t$ if*

1. *M_t is $\mathcal{M}_t$-measurable for all t*
2. *$\mathbb{E}\,|M_t| < \infty$ for all t*
3. *$\mathbb{E}(M_s|\mathcal{M}_t) = M_t$ for all $s \geq t$*

A.4 Stochastic Calculus and SDE

Stochastic Calculus starts with problems (differential equations) with random terms. For example, in process with random fluctuations, people like to consider problems with noises,

$$\frac{dX}{dt} = a(X, t) + \text{'noise'}$$

However one cannot start unless the term 'noise' is well-fined. To describe the noise mathematically, we may associate the term with a function of time W_t. A famous idealized model for the noises is the white noise where one may expect W_t to satisfy a stationary normal distribution for all t and for $t_1 \neq t_2$, W_{t_1} and W_{t_2} are independent. However, it turns out there does not exist any "reasonable" stochastic processes satisfying these property. One intuitive explanation is that the W_t here looks like the derivative of a Brownian motion while Brownian motions are nowhere differentiable.
If we rearrange the derivative in differential form and treat the term $W_t dt = dB_t$

$$dX_t = a(X_t, t)dt + W_t dt = a(X_t, t)dt + dB_t$$

Therefore, by formal notation $\int_{t_1}^{t_2} dB_t = B_{t_2} - B_{t_1}$, we have managed to deal with the noise term. However, this method fails if one wants to use variable coefficient for dB_t term. i.e. for a equation:

$$dX_t = a(X_t, t)dt + W_t dt = a(X_t, t)dt + b(X_t, t)dB_t \tag{A.1}$$

the defintion of the integral for the last term $\int_{t_1}^{t_2} b(X_t, t)dB_t$ needs to be thoroughly discussed. For example, in one dimenson, a limit of Riemann-like sum could be defined as:

$$\int_0^T b(X(t))dB(t) = \lim_{h\to 0} \sum_{n=0}^{N-1} b(X_n^*)(B_{nh+h} - B_{nh}),$$

where

$$X_n^* = (1-\alpha)X(nh) + \alpha X(nh+h)$$

with discretization $t_n = nh$, $t_N = T$. The choice of $\alpha = 0$ leads to the Itô integral. The limit here is in the sense of $\mathbb{L}^2$ in the product space of $[0, T] \times \Omega$ with weight $dt \cdot \mathbb{P}(d\omega)$.
A important property of the Itô integral is that it is a martingale. i.e. $M_t(\omega) = \int_0^t f(s, \omega)dB_s$ a martingale with respect to the filtration generated by B_t. By the martingale representation theorem [30, p. 53], the inverse is also true.
Though we have expressed X_t in terms of stochastic integrals, since the integrand depends on X_t, we haven't really solve the equation yet. One may ask a question that given an initial condition Z and a Brownian motion B_t, does there exist a process X_t satisfying the SDE (A.1) and is adapted to the filtration generated by Z and B_t? This process X_t will be called a strong solution to (A.1) since the version of the Brownian motion is given in advance. However, if one only ask for a pair of process (X_t, B_t) and they are both adapted to another filtration $\mathcal{H}_t$ (with the initial condition Z to be $\mathcal{H}_t$ measurable). Then the solution is called a weak solution. This definition is weaker since X_t may not be adapted to the filtration generated by B_t and Z any more. And there does exist the case that only weak solutions exist but not strong solutions.

A.5 The Fokker-Planck Equation and the Itô Diffusion

Consider a n-dimensional stochastic process X_t satisfying SDE

$$dX_t = a(X_t, t)dt + b(X_t, t)dB_t \tag{A.2}$$

Let $p = p(x, t|x_0, t_0)$ be the conditional probability density when $X_t = x$ and $X_{t_0} = x_0$, then p satisfies the PDE

$$\frac{\partial p}{\partial t} = -\sum_i \frac{\partial a_i(x,t)p}{\partial x_i} + \frac{1}{2}\sum_{i,j} \frac{\partial^2 [b(x,t)(b(x,t))^T p]_{ij}}{\partial x_i \partial x_j}$$

with initial condition $p(x, t_0|x_0, t_0) = \delta(x - x_0)$ where $\delta(x - x_0)$ is the delta distribution. Hence, for the same process, the two descriptions from the SDE and the PDE are to be regarded as complementary to each other.

We call the solution to (A.2) the (Itô) diffusion with a to be the drift coefficient and b or sometimes $\frac{1}{2}bb^T$ to be the diffusion coefficient. If the coefficients a and b do not explicitly depend on t, the diffusion is called time homogeneous.

A.6 Weak convergence

Given a mesurable space $(\Omega, \mathcal{F})$, if the probability measures P_n and P satisfy

$$\int_\Omega f dP_n \to \int_\Omega f dP$$

for every bounded continuous real function f on Ω. Then we say that the sequence P_n converges weakly to P and write $P_n \Rightarrow P$. In applications, people use equivalent versions of the definition to check the weak convergence. If the distributions of a sequence of random variables X_n converge weakly to the distribution of some random variable X. Then we call the sequence of random variables X_n converges to X in distribution (or weakly) and denote it as $X_n \Rightarrow X$.

Appendix B

Useful Theorems

We put the theorems used in this thesis from the literatures here.

Theorem 8 (The Arzelà-Ascoli Theorem, this statement is taken from [3])**.** *Given (S, d) a compact metric space, A set $A \subset C(S)$ is relatively compact if and only if it is composed of equicontinuous function and there exists an $M > 0$ such that*

$$\|x\|_{C(S)} \leq M, \text{ for all } x \in A$$

A is said to be composed of equicontinuous functions if and only if: $\forall \epsilon > 0$, $\forall p \in S$, there exists a $\delta > 0$, such that $\forall p' \in S$, $\forall x \in A$, $d(p, p') < \delta$ implies $|x(p) - x(p')| \leq \epsilon$.

Theorem 9 (Bakry and Emery 1985, this statement is taken from Theorem 1 in [27])**.** *Let e^{-V} be (the density of) a probability measure on $\mathbb{R}^n$, such that $D^2V \geq \lambda I_n$ where I_n is the identity matrix for dimension n and D^2 is the Hessian. Then e^{-V} satisfies a logarithmic Sobolev inequality with constant λ.*

$$H(\rho|e^{-V}) \leq \frac{1}{2\lambda} I(\rho|e^{-V})$$

where $H(\rho_1|\rho_2)$ is the relative entropy of the density ρ_1 with respect to the density ρ_2

$$H(\rho_1|\rho_2) := \int_{\mathbb{R}^d} \rho_1(x) \ln \frac{\rho_1(x)}{\rho_2(x)} dx,$$

and $I(\rho_1|\rho_2)$ is the entropy dissipation functional is defined by

$$I(\rho_1|\rho_2) := \int_{\mathbb{R}^d} \rho_1(x) \nabla \ln \frac{\rho_1(x)}{\rho_2(x)} \cdot \nabla \ln \frac{\rho_1(x)}{\rho_2(x)} dx.$$

Theorem 10 (Meyn and Tweedie 1993, this statement is taken from [20])**.** *Suppose a Markov chain has the transition kernel P given by*

$$P(x, dy) = p(x, y)dy + r(x)\delta_x(dy)$$

where δ_x is the delta distribution and $r(x)$ is the probability of staying at x. If P is irreducible and aperiodic then

$$\lim_{n\to+\infty} \|P^n(x, \cdot) - \pi(\cdot)\|_{TV} = 0$$

for π-almost all x.

The following materials are taken from [29, Ch. 2]. Consider the system,

$$dX = a(t, X)dt + \sum_{r=1}^{q} \sigma_r(t, X)dw_r(t) \tag{B.1}$$

where X, a, σ are vectors of dimension d with components X^i, a^i, σ^i. Assume that the functions $a(t, x)$ and $\sigma_r(t, x)$ satisfy a global Lipschitz condition with respect to x: for all $t \in [t_0, T]$, $x \in \mathbb{R}^d$, $y \in \mathbb{R}^d$ the following inequality holds for some positive constant K:

$$|a(t, x) - a(t, y)| + \sum_{r=1}^{q} |\sigma_r(t, x) - \sigma_r(t, y)| \le K |x - y| . \tag{B.2}$$

Here and below $|x|$ denotes the Euclidean norm of the vector x.

Definition 9. *We say that a function $f(x)$ belongs to the class $\mathbb{F}$, written as $f \in \mathbb{F}$, if we can find constants $K > 0$, $\kappa > 0$ such that for all $x \in \mathbb{R}^d$ the following inequality holds:*

$$|f(x)| \le K(1 + |x|^{\kappa}) \tag{B.3}$$

If a function $f(s, x)$ depends not only on $x \in \mathbb{R}^d$ but also a parameter $s \in S$, then we say that $f(s, x)$ belongs to $\mathbb{F}$ (with respect to the variable x) if an inequality of the type (B.3) *holds uniformly in $s \in S$.*

Along with the system (B.1), we consider the approximation

$$\bar{X}_{t,x}(t + h) = x + A(t, x, h; \xi) \tag{B.4}$$

where ξ is a random variable (in general, a vector) having moments of a sufficiently high order, and A is a vector function of dimension d. Partition the interval $[t_0, T]$ in to N equal parts with step $h = (T - t_0)/N$: $t_0 < t_1 < \ldots < t_N, t_{k+1} - t_k = h$. According to (B.1), we construct the sequence

$$\bar{X}_0 = X_0 = X(t_0), \;\; \bar{X}_{k+1} = \bar{X}_k + A(t, \bar{X}_k, h; \xi_k), \;\; k = 0, 1, \ldots, N - 1, \tag{B.5}$$

while ξ_k for $k > 0$ is independent of $\bar{X}_0, \bar{X}_1, \ldots, \bar{X}_k, \xi_0, \ldots, \xi_{k-1}$. We write $\Delta = X - x = X_{t,x}(t + h) - x, \bar{\Delta} = \bar{X} - x = \bar{X}_{t,x}(t + h) - x$. Let $X(t) = X_{t_0,x_0}(t)$ be a solution of (B.1) and $\bar{X}_{t_0,X_0}(t_k) = \bar{X}_k$ is the approximation.

Theorem 11. *Suppose that*

1. *the coefficients of equation* (B.1) *are continuous, satisfy a Lipschitz condition* (B.2) *and together with their partial derivatives with respect to* x *of order up to* $2p+2$*, inclusively, belong to* $\mathbb{F}$*;*

2. *the method* (B.4) *is such that*

$$\left|\mathbb{E}(\prod_{j=1}^{s}\Delta^{i_j}-\prod_{j=1}^{s}\bar{\Delta}^{i_j})\right|\leq K(x)h^{p+1},\quad s=1,2,\ldots,2p+1,\quad K(x)\in\mathbb{F},$$

$$\left|\mathbb{E}(\prod_{j=1}^{2p+2}\bar{\Delta}^{i_j})\right|\leq K(x)h^{p+1},\quad K(x)\in\mathbb{F};$$

3. *the function* $f(x)$ *together with its partial derivatives of order up to* $2p+2$*, inclusively, belong to* $\mathbb{F}$*;*

4. *for a sufficiently large* m *(specified below) the expectations* $\mathbb{E}\left|\bar{X}_k\right|^{2m}$ *exist and are uniformly bounded with respect to* N *and* $k=0,1,\ldots,N$*.*

Then, for all N *and all* $k=0,1,\ldots,N$ *the following inequality holds:*

$$\left|\mathbb{E}f(X_{t_0,X_0}(t_k))-\mathbb{E}f(\bar{X}_{t_0,X_0}(t_k))\right|\leq Kh^p,$$

i.e., the method (B.5) *has order of accuracy* p *in the sense of weak approximations.*

Theorem 12. *Suppose that for* $h<1$*,*

$$|\mathbb{E}A(t_k,x,h,\xi_k)|\leq K(1+|x|\,h),$$

$$|A(t_k,x,h,\xi_k)|\leq M(\xi_k)(1+|x|)h^{\frac{1}{2}}$$

where $M(\xi_k)$ *has moments of all orders. Then for every even number* $2m$ *the mathematical expectations* $\mathbb{E}\left|\bar{X}_k\right|^{2m}$ *exist and are uniformly bounded with respect to* N *and* $k=1,2,\ldots,N$*, if only* $\mathbb{E}\left|X_0\right|^{2m}$ *exists.*

Theorem 13 (Gihman and Skorohod, this statement is taken from Theorem 4 in [15])**.** *Assume the coefficients of* (B.1) *satisfy the global Lipschitz condition* (B.2) *and that* $\mathbb{E}\left|X_0\right|^{2m}$ *is finite. Then we can find a constant* C *depending only on* m*,* K *and* T *for which*

$$\mathbb{E}\left|X_t^{2m}\right|\leq\mathbb{E}(1+\left|X_0\right|^{2m})\exp(Ct)$$