Scaling Limits of Active Particles

Sven Holtrop, 4933125

Bachelor Thesis,
Applied Mathematics and Physics,
Delft University Of Technology

Delft, 5 Juli 2021
Supervisor:
Prof. dr. F.H.J. Redig
Dr. J.M. Thijssen

Abstract

In this thesis, the diffusive limit of active particle motion in $\mathbb{R}^d$ is studied via a technique based on homogenisation. Thereafter, this study is extended to active particle motion on a Riemannian manifold.

Furthermore, as an application of active particle motion, a connection is made with the Dirac equation. On the basis of this connection, a Monte Carlo method is developed to find the ground state of a Dirac equation with static potential. The core idea of this method is based on the Diffusion Monte Carlo method for the Schrödinger equation.

Contents

1 Introduction

We, as humans, have a plethora of tendencies. One such tendency is thinking in terms of sequences of events. Even when there does not seem to be a direct causal link within a sequence, humans will still try to describe it in terms of what they think likely to happen next. In mathematics, we call such a probabilistic sequence of events a stochastic process. In this Bachelor thesis, we will focus on one specific stochastic process, namely the active particle process.

In the simplest terms, this process consists of a particle, which on the one hand moves around randomly, but on the other hand has internal state, which dictates a preferred direction of motion. However, this internal state is subject to a random process itself. One physical example of such particles are motor proteins, which on the one hand, are subject to random collisions from surrounding molecules, but on the other hand, try to move in a specific direction based on some internal chemical state. One of the main questions of interest is what the long term behaviour of such a particle is. This can be studied through a tool called the diffusive limit. Recently in 2021, this limit was calculated for $\mathbb{R}^d$-valued active particles in [1].

Now, another setting in which stochastic processes can be studied apart from $\mathbb{R}^d$ is Riemannian manifolds; especially smooth surfaces are of interest. Some of the earlier work, which explores stochastic processes in such a setting, would be [2]. However, more recently the large deviations behaviour of process on manifolds was studied in [3] and simulations for systems which resemble active particles on manifolds were done in [4, 5]. Stochastic processes on manifolds differ from $\mathbb{R}^d$ due to the effects of curvature in a manifold. For a physical example of an active particle process on a manifold, we return to biology, where we can consider active particles on a cell membrane. This membrane can be viewed as 2-dimensional embedded submanifold of $\mathbb{R}^3$ and so, we see that the diffusion of the active particles might depend on the curvature of the membrane.

With this interest in mind, this Bachelor thesis aims to introduce a new way of calculating the diffusive limit of active particles in $\mathbb{R}^d$ – specifically through homogenisation – and to then extend this result to Riemannian manifolds. Apart from this type of limit, another type is also considered, which converges to a telegrapher's process. Furthermore, an application of active particles in physics is highlighted by connecting an active particle process to the Dirac equation – a quantum mechanical equation for electrons in the relativistic regime – and then using this connection, a numerical method for solving the 1-dimensional Dirac equation is developed. Specifically, this is a Monte Carlo method, which approximates the solution on the basis of simulating active particle trajectories.

This thesis starts out with two chapters introducing the theory of stochastic processes. In the mathematical background chapter, we introduce some basic material about Markov processes and their associated martingales. Roughly speaking, we say that a stochastic process is Markov if the distribution of future states only depends on its present state and not on the further past. This introductory chapter might be skipped by experienced readers. However, on the other hand, if this chapter is too advanced, appendix A might aid in covering some of the prerequisites. Thereafter, there is a chapter on random walks and Brown motion in $\mathbb{R}^d$, arguably the most important stochastic processes that exist. In this chapter, diffusive scaling is also introduced.

Then, two chapters on active particles follow. Firstly, a chapter on $\mathbb{R}^d$-valued active

particles in which all the basic notions are introduced and the proof that is to be extended, is given. Thereafter, the previous chapter is repeated, only now the active particle process is manifold-valued.

Finally, three chapters on applications of stochastic process in physics follow. Firstly, the Schrödinger equation and its link with Brownian motion is studied. This link can be used to create the Diffusion Monte Carlo algorithm, which finds the ground state of the Schrödinger equation. Thereafter, in the next chapter, the Dirac equation is introduced and it is linked to active particles. The final chapter introduces a new algorithm, which is based on the same ideas as the Diffusion Monte Carlo algorithm, to find the ground state of the Dirac equation.

2 Mathematical background

Active random walks are a type of stochastic process. Thus before we can properly examine them, we first have to develop a more general framework for understanding stochastic processes. We will start this chapter by giving a definition for a stochastic process. We will then continue introducing time-homogeneous Markov processes and a restricted variant thereof called Feller processes. Finally, Martingales – a tool used extensively throughout probability theory – is introduced.

Definition 2.1. *Let $(\Omega, \mathscr{A}, \mu)$ be a probability space, $(E, \mathscr{E})$ a measurable space and T a totally ordered set. We then call a map $X : T \times \Omega \to E$ a stochastic process if it is measurable in the product σ-algebra $\mathscr{B}(T) \otimes \mathscr{A}$. We also write $X_t = X(t, \cdot) : \Omega \to E$.*

In the above definition, we will always take T to be either $\mathbb{R}_{\geq 0}$ or $\mathbb{N}$; and call it continuous and discrete time respectively. Furthermore, if we can define an expectation $\mathbb{E}$, a stochastic process X can be viewed as a sequence of random variables X_t.

Before we start, some notation: we use $L^p(E, \mathcal{E}, \mu)$ to denote the space of functions $f : E \to \mathbb{R}$ with $\mathbb{E}[|f|] < \infty$ and $L(E, \mathcal{E})$ for the space of measurable functions $f : E \to \mathbb{R}$. $C(E)$ denotes the continuous functions $f : E \to \mathbb{R}$. The addition of a subscript b indicates the functions are bounded. Furthermore,if E is a locally compact metric space, a subscript 0 indicates that the functions go to 0 at infinity. In all of the above cases, we might leave out specifiers for the base space if there is only one clear candidate. So if we are talking about some space E and then say $f \in C_b$, we mean that is a continuous bounded function from E into $\mathbb{R}$.

2.1 Markov Processes

Firstly, this section is based on [6, 7]. Informally speaking, a Markov process is a process without memory. However, what does it entail for a process to be memoryless? Suppose that we have some $(E, \mathscr{E})$-valued process $(X_t)_{t \in T}$ and that we are currently at $s \in T$. We then say that the process is memoryless if some future state $X_{t'}$ only depends on the current state X_s; in other words, conditioned on the present, the future does not depend on the past.

To make this more precise, we will first introduce the natural filtration with respect to X, which is defined as $(\mathscr{F}_t^X)_{t \in T}$ with $\mathscr{F}_t^X = \sigma(\bigcup\{X_s^{-1}(\mathscr{E}) : s \leq t\})$. We can now define Markov processes.

Definition 2.2. *Let $X = (X_t)_{t \in T}$ be a stochastic process in $(E, \mathscr{E})$. Then X is called a Markov process if for all $0 < s \leq t$ and for every measurable bounded function $f : E \to \mathbb{R}$,*

$$\mathbb{E}[f(X_t)|\mathscr{F}_s^X] = \mathbb{E}[f(X_t)|X_s].$$

The above definition, although pleasing, is not very tangible when it comes to defining one's own Markov processes. Firstly, we will assume that our Markov processes are time-homogeneous. This means that transitions between states only depend on the state itself and not on the time of transitioning. Then, the idea of a transition kernel naturally arises.

Definition 2.3. *Let $(E, \mathscr{E})$ be a measurable space. Then, the map $S : E \times \mathscr{E} \to [0,1]$ is called a* Markovian transition kernel *on E if*

(i) *for all $x \in E$, the map $A \to S(x, A)$ is a probability measure on $(E, \mathscr{E})$;*

(ii) *for all $A \in \mathscr{E}$, the map $x \to S(x, A)$ is measurable.*

Thus, the transition kernel is way of specifying how a process transitions when in a certain state $x \in E$. Now, a typical way for a measure to act on a function is by means of an integral or in the context of probability theory an expectation. So for $f \in L_b(E, \mathscr{E})$, we might define Sf as the function:

$$Sf(x) := \int_E f(y)\, S(x, \mathrm{d}y),$$

which is the expectation of f after a transition form x according to S. Note that Sf is measurable and in addition S is a bounded linear operator with $\|S\| = 1$.

Proof. Firstly, note that as integration is linear, S is a linear operator. Now, let $f \in L_b(E, \mathscr{E})$. Then suppose $f = \mathbf{1}_A$ with $A \in \mathscr{E}$. Then $Sf(x) = \int_E \mathbf{1}_A(y)\, S(x, \mathrm{d}y) = S(x, A)$, which is measurable by condition (ii) from definition 2.3. Now, suppose $f = \sum_k a_k \mathbf{1}_{A_k}$ is a simple function. Then $Sf = \sum_k a_k S(x, A_k)$ is measurable, since it is a linear combination of measurable functions. Furthermore, if f is a positive function, we can find a sequence of simple functions f_n increasing to f. By the monotone convergence theorem, we have $\lim Sf_n = Sf$ pointwise. As the Sf_n are simple, they are measurable. Then, since Sf is a pointwise limit of measurable functions, Sf is measurable itself. For general $f \in L_b(E, \mathscr{E})$, measurability follows from the linear decomposition $f = f^+ - f^-$ with $f^+, f^- \geq 0$.

Finally, as f is bounded, $|Sf(x)| \leq \int_E |f(y)|\, S(x, \mathrm{d}y) \leq \|f\| \int_E S(x, \mathrm{d}y) = \|f\| \leq \infty$. Taking $f = 1$, we find $S1 = \int_E S(x, \mathrm{d}y) = 1$. So S is a bounded linear operator with $\|S\| = 1$. □

Now, we would like to associate transition kernels with certain times $t \in T$. This brings us to the following definition.

Definition 2.4. *A collection $(S_t)_{t \geq 0}$ of transition kernels on E is called* transition semigroup *if*

(i) *for all $x \in E$, $S_0(x, \mathrm{d}y) = \delta_x(\mathrm{d}y)$;*

(ii) *(Chapman-Kolmogorov equation) for all $s, t \geq 0$ and $A \in \mathscr{E}$,*

$$S_{s+t}(x, A) = \int_E S_s(y, A)\, S_t(x, \mathrm{d}y);$$

(iii) *for all $A \in \mathscr{E}$, the map $t, x \mapsto S_t(x, A)$ is measurable with respect to the product σ-algebra $\mathscr{B}(T) \otimes \mathscr{E}$.*

Now, suppose we have a time-homogeneous Markov process $(X_t)_{t\in T}$. Then, $S_t(x,A) = \mu(X_t \in A|X_0 = x)$ defines a semigroup and $S_t f(x) = \mathbb{E}[f(X_t)|X_0 = x]$. Moreover by time-homogeneity, $S_t f(x) = \mathbb{E}[f(X_{s+t})|X_s = x]$ and thus we see that $\mathbb{E}[f(X_t)|\mathscr{F}_s^X] = \mathbb{E}[f(X_{s+t})|X_s] = S_t f(X_s)$. Note that $S_t f(X_s) \neq \mathbb{E}[f(X_{s+t})|X_s = X_s] = \mathbb{E}[f(X_{s+t})]$. Instead, we associate to each fibre $X_s^{-1}(\{x\})$ the value $\mathbb{E}[f(X_{s+t})|X_s = x]$ culminating in $\mathbb{E}[f(X_{s+t})|X_s]$.

To see S_t is a semigroup, firstly note that $S_0(x, \mathrm{d}y) = \mu(X_0 \in \mathrm{d}y\,|X_0 = x) = \delta_x(\mathrm{d}y)$. Furthermore,

$$\begin{aligned}
S_{t+s}(x,A) &= \mu(X_{s+t} \in A|X_0 = x) \\
&= \int_E \mu(X_{s+t} \in A|X_t = y, X_0 = x)\, \mu(X_t \in \mathrm{d}y\,|X_0 = x) \\
&\overset{*}{=} \int_E \mu(X_{s+t} \in A|X_t = y)\, S_t(x, \mathrm{d}y) \\
&\overset{**}{=} \int_E \mu(X_s \in A|X_0 = y)\, S_t(x, \mathrm{d}y) \\
&= \int_E S_s(y,A)\, S_t(x, \mathrm{d}y) = S_t S_s(x,A),
\end{aligned}$$

where at $*$, the Markov property was used and at $**$, the time-homogeneity of the process and the identity $S_t f(x) = \mathbb{E}[f(X_t)|X_0 = x]$ were used. The final property follows from the measurability of the process X with respect to $\mathscr{B}(T) \otimes \mathscr{A}$.

But what is the point of these semigroups? The beauty of semigroups is that there exists a unique association between them and time-homogeneous Markov processes. So if we want to prove two processes are the same, it suffices to show they have the same semigroups.

Theorem 2.5. *Two Markov processes are the same if and only if they have the same semigroups.*

Example 1. As a first example, we will consider Markov processes with a countable state space E in discrete time. In this case, we have a transition probability for transitioning from state x to state y in one time step, denoted $p(x,y)$. As a transition must always occur, we find that $\sum_{y\in E} p(x,y) = 1$. We can then specify the potentially infinite matrix $P_{xy} = p(x,y)$. In addition, we can associate a distribution μ with the vector $\vec{\mu}_x = \mu(x)$.

Suppose we now have some initial distribution μ^0. Then, the distribution at the next time step is given by $\mu^1(x) = \sum_{y\in E} \mu^0(y)p(y,x) = \sum_{y\in E} \vec{\mu}^0_y P_{yx}$ or simply $\vec{\mu}^1 = P^T\vec{\mu}^0$. Thus, we see that $\vec{\mu}^n$ is given by $(P^n)^T\vec{\mu}^0$.

By identifying a state x with the vector $\vec{\delta}_x$ and a set $A \subseteq \mathscr{E}$ with the vector $\vec{A}$, where $\vec{A}_x = 1$ if $x \in A$ and $\vec{A}_x = 0$ otherwise, we see that $P_n(x,A) := \vec{\delta}_x^T P^n \vec{A}$ defines a semigroup, where we essentially transport the distribution $\vec{\delta}_x$ into the future using P^n. So, in the case of Markov processes with a countable state space in discrete time, we have that the powers of the transition matrix P specify the semigroup in its entirety. Thus, the matrix P characterises the Markov process.

Let us now have a look at the action of P on a function f. For a countable state space, we can associate a function $f : E \to \mathbb{R}$ with the column vector $\vec{f}_x = f(x)$, so that

$f(x) = \vec{\delta}_x^T \vec{f}$. We then have that $P_n f(x) = \mathbb{E}[f(X_n)|X_0 = x] = \sum_{y\in E} P^n(x, \{y\})f(y) = \sum_{y\in E} \vec{\delta}_x^T P^n \vec{\delta}_y f(y) = \vec{\delta}_x^T P^n \vec{f}$. Thus with some abuse of notation, we can say $P_n f = P^n \vec{f}$.

To make this even more concrete, we can consider a discrete time random walk on $\mathbb{Z}$. In a random walk, we start out at $n = 0$ with a particle at the origin. Then, there is a probability of $\frac{1}{2}$ of transitioning to -1 or similarly to $+1$. From this new position, we repeat the same procedure. Thus, for a particle in position x, there is a probability of $\frac{1}{2}$ of transitioning to $x-1$ or similarly to $x+1$. Equivalently, we can say the particle jumps one step to the right or to the left with equal probability. Based on this, we can specify a transition matrix, namely $P_{xy} = \frac{1}{2}\delta_{x-1}^y + \frac{1}{2}\delta_{x+1}^y$. △

That we can characterise an entire Markov process by a single matrix – as above – is an extremely useful property. For a Markov process with an uncountable state space in discrete time, we can do something similar to the above procedure by chaining a single time step transition function. However, in the continuous time scenario, this is not as simple and we need some more considerations.

2.2 Feller Processes

Let the first remark be that this section is largely based on the same material as the previous section. In addition, the book foundations of probability theory by Olav Kallenberg [8] is used.

Continuous time processes are significantly more difficult to study than their discrete time counterparts. Firstly, are continuous time processes measurable with respect to $\mathscr{B}(\mathbb{R}_{\geq 0}) \otimes \mathscr{E}$? For a discrete time process X, we simply require that X_n is measurable for all $n \in \mathbb{N}$. However, for continuous time processes, this is not sufficient. There exist several classes of measurable continuous time processes. We will restrict ourselves to focusing on the so called càdlàg processes. This is an acronym standing for *"continue à droite, limite à gauche"*. This means realisations of a process, which are functions of time, are continuous to the right of every point and have a left-handed limit at every point. Or more colloquially, any discontinuities in the graph of a realisation occur at the left side of the discontinuity point. An example of such a curve can be seen in figure 1. This property guarantees that the process is measurable.

Secondly, we cannot simply characterise the entire process with a single transition function for a fixed time interval anymore, as such a function could never cover the entirety of $\mathbb{R}_{\geq 0}$. However, for a special class of processes called Feller processes, we can find an object that characterises the semigroup and thereby the associated Markov process.

Definition 2.6. *We call a transition semigroup $(S_t)_{t\geq 0}$ on a locally compact metric space E a Feller semigroup if for all $f \in C_0$,*

(i) $S_t f \in C_0$

(ii) $\|S_t f - f\| \to 0$ *as* $t \to 0$,

where C_0 denotes the continuous functions that vanish at infinity.

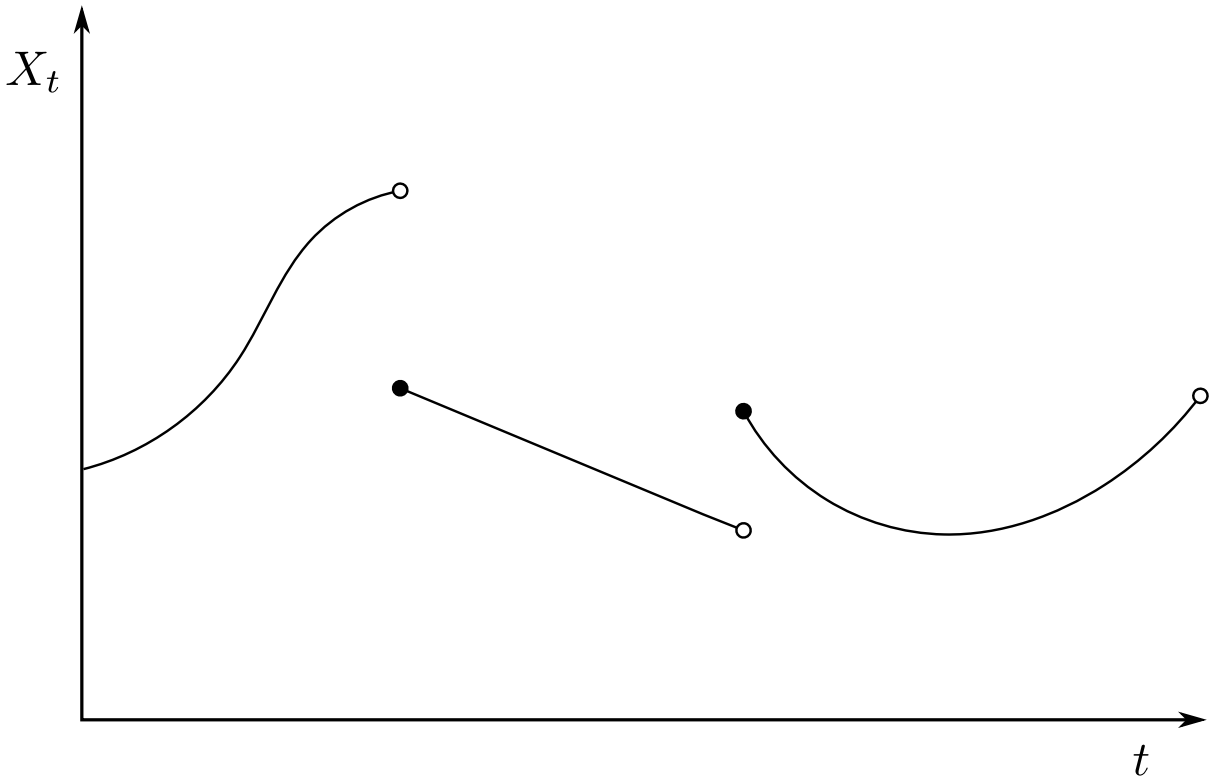

Figure 1: An example of a càdlàg process, where an open dot denotes that no point is included here in the curve and a closed dot denotes that a point is included here.

A Markov process with a Feller semigroup is called a Feller process. We would like to note that the second condition is equivalent to $S_t f$ being càdlàg as a function of time. Firstly, (ii) follows immediately from being $S_t f$ right continuous as function of time. On the other hand, if (ii) holds, then $\lim_{h\downarrow 0} \|S_{t+h}f - S_t f\| = \lim_{h\downarrow 0} \|S_t\| \|S_h f - f\| = 0$, where we used the Chapman-Kolmogorov equation in the second to last equality.

So, we then have that our semigroup has the following properties $S_0 = I$, $S_{s+t} = S_s S_t$ and it is somewhat continuous as a function of time. These properties might remind one of an exponential function $f(t) = e^{at}$ which is characterised by $f(0) = 1$, $f(s+t) = f(s)f(t)$ and f being continuous [9]. This reminiscence can be made rigorous through the idea of a strongly continuous semigroup. This is a generalisation of the exponential functions to operators. And just as e^{at} is characterised by a single value a, a strongly continuous semigroup is characterised by a single operator.

Definition 2.7. *Let $(S_t)_{t\geq 0}$ be a semigroup. We call $D(L)$ the domain of the generator and define it as*

$$D(L) = \left\{ f \in C_0(E) : \lim_{t\to 0} \frac{S_t f - f}{t} \text{ exists} \right\},$$

where the limit is with respect to the supremum norm. Moreover, we define the generator of $(S_t)_{t\geq 0}$ as the linear operator $L : D(L) \to C_0(E)$ with

$$Lf = \lim_{t\to 0} \frac{S_t f - f}{t}.$$

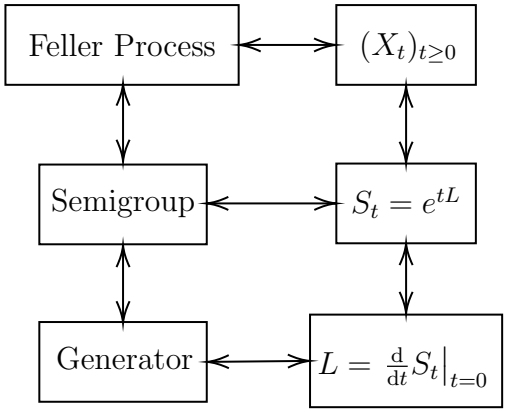

Figure 2: A diagram giving an overview of Feller processes. The left column contains the names of the objects. The right column denotes the attached symbols and their relations.

The above definition is an exact parallel to $a = \frac{\mathrm{d}}{\mathrm{d}t} e^{at}\big|_{t=0}$. However, this is not the only way of characterising e^{at}. In this spirit, we have the next theorem for Feller semigroups.

Theorem 2.8. *Two Feller semigroups are the same if and only if their generators and domains are identical. In addition, we have that*

(i) $S_t = e^{tL} = \sum_{n=0}^{\infty} \frac{(tL)^n}{n!}$;

(ii) $S_t f$ *is the unique solution to differential equation* $\frac{\mathrm{d}}{\mathrm{d}t} g_t = L g_t$ *with the initial condition* $g_0 = f$.

The above shows that we can characterise a Feller process X_t solely through its associated generator, since the semigroup fully determines the stochastic process X_t and the semigroup is fully characterised by its generator L. This web of relationships is presented diagrammatically in figure 2.

Before we will list some useful properties of generators, we will first give a very quick overview of *invariant* or *stationary* measures. Suppose we have a Feller process with semigroup $(S_t)_{t\geq 0}$ and generator L. We then call a measure μ invariant if $\int S_t f \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$ for all $f \in C_0(E)$ and $t \geq 0$. Moreover, a measure μ is invariant if and only if $\int Lf \,\mathrm{d}\mu = 0$ for all $f \in D(L)$.

Proposition 2.9. *Let* $(S_t)_{t\geq 0}$ *be a Feller semigroup with generator* L*, let* $f \in D(L)$*, let* $c \in \mathbb{R}$ *be a constant and define* $L^{-1} f = -\int_0^\infty S_t f \,\mathrm{d}t$*, then it holds that*

(i) $L(c) = 0$;

(ii) $S_t L f = L S_t f$;

(iii) *if there exists a unique stationary measure* μ *and* $K, C \in \mathbb{R}_{>0}$ *such that* $\| S_t f - \int f \,\mathrm{d}\mu \|_{L^2(\mu)} \leq K e^{-Ct}$*, then the linear operator* $L^{-1} f = -\int_0^\infty S_t f \,\mathrm{d}t$ *exists and* $L^{-1} L f = L L^{-1} f = f$.

Proof. Firstly, note that $S_t(c) = \int_E c\, S_t(x, \mathrm{d}y) = c$. Thus, $L(c) = \lim_{t\to 0} \frac{S_t c - c}{t} = \lim_{t\to 0} \frac{c-c}{t} = 0$.

Secondly, note that S_t is a bounded linear operator, so $S_t L f = S_t \lim_{s\to 0} \frac{S_s f - f}{s} = \lim_{s\to 0} \frac{S_{s+t} f - S_t f}{s} = \lim_{s\to 0} \frac{S_s f - I}{s} S_t f = L S_t f$.

Lastly, by theorem 2.8 (ii), $L^{-1} L f = -\int_0^\infty S_t L f \, \mathrm{d}t = -\int_0^\infty L S_t f \, \mathrm{d}t = -\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t} S_t f \, \mathrm{d}t = \lim_{t\to\infty}(S_0 f - S_t f) = f - \lim_{t\to\infty} S_t f = f$. In addition, for a invariant measure μ and $f \in D(L)$, it holds that $\int f \, \mathrm{d}\mu = 0$, so $\|S_t f\|_{L^2(\mu)} = \|S_t f - \int f \, \mathrm{d}\mu \,\|_{L^2(\mu)} \leq K e^{-Ct}$. So, $S_t f$ is integrable and thus, L^{-1} exists for all $f \in D(L)$. Moreover, we find using Fubini's theorem at $*$ and the fact that Lf is integrable (all C_0 functions are bounded), $L L^{-1} f = -L \int_0^\infty S_t f \, \mathrm{d}t = -\lim_{s\to 0} \frac{S_s \int_0^\infty S_t f \mathrm{d}t - \int_0^\infty S_t f \mathrm{d}t}{s} \overset{*}{=} -\lim_{s\to 0} \int_0^\infty \frac{S_s(S_t f) - S_t f}{s} \, \mathrm{d}t \overset{DCT}{=} -\int_0^\infty L S_t f \, \mathrm{d}t = f$. □

In particular, any finite state irreducible aperiodic Markov chain satisfies the assumptions of proposition 2.9 (ii). Another way to check if an inverse exists, is to explicitly check that $\int_0^\infty S_t f \, \mathrm{d}t$ exists.

Example 2. Let us now extend the example of a discrete state space in discrete time to the continuous time case. This is called a jump process. In a jump process, we wait for an exponential time in some state $s \in E$, before jumping to a next state $s' \in E$, where this transition is dictated by the single time step transition from the original discrete time process. The exponential distribution is in fact the only possibility here, as it is the only distribution which satisfies the Markov property [9]. From the same source, we know that the cumulatie number of jumps N_t then has a Poisson distribution with mean λt and that $P(N_{t+\tau} = n+1 | N_t = n) = \lambda\tau + o(\tau)$, $P(N_{t+\tau} = n | N_t = n) = 1 - \lambda\tau + o(\tau)$ and $P(N_{t+\tau} = n+k | N_t = n) = o(\tau)$ for $k > 1$. Here, $o(f(x))$ denotes that this part of the expression goes to 0 as $x \to 0$ after dividing by $f(x)$

Suppose we have a discrete time process $(P_n)_{n\geq 0}$ with the single time step transition kernel $P_1 = P$. We will now make this process continuous by jumping to a new state at rate κ. We say the continuous process has semigroup S_t. Now, we find that $S_t f(x) = \sum_{n=0}^\infty \mathbb{P}(N_t = n) P_n f(x) = P_0 f(x)(1 - \kappa t) + P_1 f(x)\kappa t + o(t)$. So $Lf(x) = \lim_{t\to 0} \frac{S_t f(x) - f(x)}{t} = \lim_{t\to 0} \frac{\lambda t (P_1 f(x) - P_0 f(x)) + o(t)}{t} = \lambda(Pf(x) - f(x))$. Thus, $Lf = \lambda(Pf - f)$.

To make this more concrete, we will now take the process (P_n) to be the symmetric random walk on $\mathbb{Z}$. Then $P_{xy} = \frac{1}{2}\delta^y_{x-1} + \frac{1}{2}\delta^y_{x+1}$. So we see that $Lf(x) = \lambda(Pf(x) - f(x)) = \frac{\lambda}{2}(\sum_{y\in\mathbb{Z}}(\delta^y_{x+1} f + \delta^y_{x-1} f)(x) - 2f(x)) = \frac{\lambda}{2}(f(x+1) + f(x-1) - 2f(x))$.

Finally, suppose we have a finite set of one time step transition kernels P_k on the state space E with $k \in \{1, ..., K\}$. We can then construct a single continuous time process by letting the process jump independently at rate κ_k according to the transition kernel P_k for all $k \in \{1, ..., K\}$. This is in fact equivalent to a process which jumps with rate $\kappa = \sum \kappa_k$ according to the one time step transition kernel $P = \frac{1}{\kappa}\sum \kappa_k P_k$. We then find that the generator is given by $Lf = \kappa(Pf - f) = \sum \kappa_k P_k f - \kappa f = \sum \kappa_k (P_k f - f)$. △

All this has brought us to the final theorem of this section. Since this bachelor thesis focuses on scaling limits of processes, we would like to link the convergence of generators to the convergence of processes. The next quite remarkable theorem from [8] does this.

Theorem 2.10. *Let $X, (X^n)_{n\geq 0}$ be Feller processes with semigroups $S_t, (S_{n,t})_{n\geq 0}$ and generators $L, (L_n)_{n\geq 0}$. Then the following are equivalent:*

(i) *for every* $f \in D(L)$*, there exists a sequence* $f_n \in D(L_n)$ *such that* $f_n \to f$ *and* $L_n f_n \to Lf$ *with respect to* $\|\cdot\|$*;*

(ii) $S_{n,t} \to S_t$ *in the supremum norm* $\|S\| = \sup_{\|f\|\neq 0} \frac{\|Sf\|}{\|f\|}$*;*

(iii) $X^n \xrightarrow{d} X \implies$ *convergence in the space of sample paths.*

As a final note, in some cases, we might want to apply the generator to a function which is not contained in its domain. In particular, we are interested in the first and second moment of random variables. Even though these functions are not part of the domain of the generator, since they do not vanish at infinity, there are no problems when applying the generator to these functions, as long as we can evaluate the generator and the moments are integrable. In this case, it is possible to extend the definition of the generator through something called the full generator.

2.3 Martingales

Martingales are a tool used extensively throughout probability, because a wide variety of useful results exist regarding the behaviour of martingales. This subsection will look at an extremely narrow selection of all the existing material. For more information, see [8, 7], which also form the basis of this subsection.

In the first subsection of this chapter, we have already defined the natural filtration of a stochastic process. This is actually just an example of a filtration.

Definition 2.11. *Let* $(\Omega, \mathscr{F}, \mu)$ *be a probability space. A collection of sub-*σ*-algebras* $(\mathscr{F}_t)_{0\leq t\leq\infty}$ *is called a filtration if* $\mathscr{F}_s \subseteq \mathscr{F}_t$ *for all* $0 \leq s \leq t \leq \infty$*. We call a stochastic process* $(X_t)_{t\geq 0}$ $\mathscr{F}_t$*-adapted if* X_t *is* $\mathscr{F}_t$*-measurable.*

The idea that σ-algebras represent information, leads to the idea that a filtration is a probabilistic way of information being revealed with time. For example, if you throw a die every second, the filtration increasing through time represents the fact that we know the outcome of more dice rolls as time progresses.

Definition 2.12. *Let* $(\mathscr{F}_t)_{t\geq 0}$ *be a filtration and* $(X_t)_{t\geq 0}$ *an* $\mathscr{F}_t$*-adapted process. Then we call* X_t *a martingale if for all* $t \geq 0$*,*

(i) $X_t \in L^1$*;*

(ii) *for all* $0 \leq s \leq t$*,* $\mathbb{E}[X_t|\mathscr{F}_s] = X_s$*.*

An important property of a martingale is its quadratic variation. This is a process associated with a martingale that measures how variation accumulates throughout the process. In addition, we might define an analogue to covariance for processes called the cross variation. The following definition – and at the same time theorem – specifies what is exactly meant by this.

Definition 2.13. *Let M, N be a continuous-time $\mathbb{R}$-valued martingales. Then, there exists a unique quadratic variation process denoted $\langle M \rangle$ such that $M^2 - \langle M \rangle$ is a martingale. Moreover, for all $t > 0$ and any collection of increasing sequences $0 = t_0^n < t_1^n < ... < t_{p_n}^n = t$ such that $\sup_{i \neq j} |t_i^n - t_j^n| \to 0$ as $n \to \infty$, we have that*

$$\langle M \rangle_t = \lim_{n \to \infty} \sum_{k=1}^{p_n} (M_{t_k^n} - M_{t_{k-1}^n})^2.$$

Furthermore, we define the cross variation process $\langle M, N \rangle = \frac{1}{4}(\langle M + N \rangle - \langle M - N \rangle)$, which is the unique process such that $MN - \langle M, N \rangle$ is a martingale.

Note that one can think of $\langle M, N \rangle$ as an inner product. In that sense, we also have $\langle M \rangle = \langle M, M \rangle$ and the definition of cross variation is simply the usual polarisation relation between the norm and inner product. The following two theorems show how to construct a martingale from a Feller process and how to calculate the quadratic variation of this martingale.

Theorem 2.14. *Let $(X_t)_{t \geq 0}$ be a Feller process with generator L and $f \in D(L)$. Then, we find that M^f is a martingale for all $f \in D(L)$ with respect to the natural filtration $(\mathscr{F}_t^X)_{t \geq 0}$, where*

$$M_t^f = f(X_t) - f(X_0) - \int_0^t L(f)(X_s) \, \mathrm{d}s \, .$$

Proof. Firstly, note M^f is clearly $\mathscr{F}_t^X$-adapted and that $\mathbb{E}[| \int_0^t L(f)(X_s) \, \mathrm{d}s \, |] \leq \|L(f)\| t < \infty$. As we also have $f \in C_0(E)$, we find that $M_t^f \in L_1$ for all $t \geq 0$.

Secondly, suppose we have $0 \leq s < t$, then

$$\begin{aligned}
\mathbb{E}[M_t^f - M_s^f | \mathscr{F}_s^X] &= \mathbb{E}\left[f(X_t) - f(X_s) - \int_s^t L(f)(X_u) \, \mathrm{d}u \, | \mathscr{F}_s^X \right] \\
&\overset{*}{=} \mathbb{E}[f(X_t) | X_s] - f(X_s) - \int_s^t \mathbb{E}[L(f)(X_u) | X_s] \, \mathrm{d}u \\
&\overset{**}{=} S_{t-s} f(X_s) - f(X_s) - \int_0^{t-s} S_u L(f)(X_s) \, \mathrm{d}u \\
&\overset{***}{=} S_{t-s} f(X_s) - f(X_s) - \int_0^{t-s} \frac{\mathrm{d}}{\mathrm{d}u} S_u(f)(X_s) \, \mathrm{d}u \\
&= S_{t-s} f(X_s) - f(X_s) - S_{t-s} f(X_s) - S_0 f(X_s) = 0,
\end{aligned}$$

where at *, it is used that X_t is a Markov process; at **, Fubini's theorem is applied and it is used that S_{t-s} can be thought of as propagating X_s to X_t; and at ***, theorem 2.8 (ii) is used. So since $\mathbb{E}[M_t^f - M_s^f | \mathscr{F}_s^X] = 0$, we have $\mathbb{E}[M_t^f | \mathscr{F}_s^X] = \mathbb{E}[M_s^f | \mathscr{F}_s^X] = M_s^f$. □

Note that M^f is linear in f, since integration and the generator are linear operators. Thus, $M^{\alpha f + \beta g} = \alpha M^f + \beta M^g$.

Theorem 2.15. *Let $(X_t)_{t \geq 0}$ be a Feller process with generator L. We then find that*

$$\langle M^f \rangle_t = \int_0^t \left[L(f^2)(X_s) - 2 f(X_s) L(f)(X_s) \right] \mathrm{d}s \, .$$

Proof. See section 8.1 of [10]. □

Then, we find for the covariance process

$$\begin{aligned}
\langle M^f, M^g\rangle_t &= \frac{1}{4}(\langle M^f + M^g\rangle - \langle M^f - M^g\rangle) \\
&= \frac{1}{4}(\langle M^{f+g}\rangle - \langle M^{f-g}\rangle) \\
&= \frac{1}{4}\int_0^t \big[L(f^2+2fg+g^2)(X_s) - 2(f+g)(X_s)L(f+g)(X_s) \\
&- L(f^2-2fg+g^2)(X_s) + 2(f-g)(X_s)L(f-g)(X_s)\big]\,\mathrm{d}s \\
&= \frac{1}{4}\int_0^t [4L(fg)(X_s) - 4f(X_s)L(g)(X_s) - 4g(X_s)L(f)(X_s)]\,\mathrm{d}s \\
&= \int_0^t [L(fg)(X_s) - f(X_s)L(g)(X_s) - g(X_s)L(f)(X_s)]\,\mathrm{d}s\,.
\end{aligned}$$

Example 3. We will continue with the continuous time random walk on $\mathbb{Z}$ with $Lf(x) = \frac{\lambda}{2}(f(x+1) + f(x-1) - 2f(x))$. We then see that when we take $f(x) = x$, $M_t^f = X_t - X_0 - \frac{\lambda}{2}\int_0^t [X_s + 1 + X_s - 1 - 2X_s]\,\mathrm{d}s = X_t$ i.e. X_t is a martingale. This process then has quadratic variation $\langle M^f\rangle_t = \frac{1}{2}\int_0^t [(X_s+1)^2 + (X_s-1)^2 - 2X_s^2 - 2X_s(X_s+1+X_s - 1-2X_s)]\,\mathrm{d}s = \frac{1}{2}\int_0^t [2X_s + 1 - 2X_s + 1]\,\mathrm{d}s = \int_0^t \mathrm{d}s = t$ i.e. $X_t^2 - t$ is a martingale. △

3 Random walks and Brownian motion

In this chapter, which is largely based on [6], we will explore Brownian motion, random walks and the relation between these two types of processes. In addition, we will introduce the diffusive scaling limit. As a motivation for why this is interesting, we start by recalling the central limit theorem.

Theorem 3.1 (Central Limit Theorem)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. L^2 random variables. Take $\mu = \mathbb{E}[X_1]$ and $\sigma^2 = \mathrm{Var}(X_1)$. Then,*

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu) \xrightarrow{d} \mathcal{N}(0,\sigma^2) \text{ as } n \to \infty,$$

where $\mathcal{N}(a, b^2)$ denotes the normal distribution with mean a and variance b^2.

The central limit theorem is an example of a scaling limit for random variables and one of the most widely used theorems in statistics and probability. The idea behind the factor $\frac{1}{\sqrt{n}}$ is to keep the variance of the sum of random variables constant and equal to σ^2. Moreover, notice that if X_i is a sequence of i.i.d. random variables and $X_0 \sim \mathcal{N}(0,\sigma^2)$, then $\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu) \overset{d}{=} \mathcal{N}(0,\sigma^2)$ for all $n \in \mathbb{N}$. So, we observe that the normal distribution is invariant under diffusive scaling.

This scaling as done in the central limit theorem is in fact the simplest form of so called diffusive scaling. For stochastic processes, we can also define such a diffusive scaling. We then find that Brownian motion fulfils a role similar to the normal distribution, as Brownian motion is invariant under this scaling.

3.1 Random Walks

As an example, we have already considered random walks on $\mathbb{Z}$. We will now define a general random walk in $\mathbb{R}^d$. Firstly, we will consider a discrete time random walk. Suppose we start out with some particle at the origin. Then, at each time step, the particle will jump along a vector to a new position. We take this vector to be an $\mathbb{R}^d$-valued L^2 random variable Z with mean 0. So essentially, if we have particle located at x, the particle will transition to $x + Z$ in the next time step. The path traced out by the particle is the corresponding random walk process $(X_n)_{n\geq 0}$. Then, the single time step transition kernel of X_n is given by

$$P(x, A) = \int_{\mathbb{R}^d} \mathbf{1}_A(x + Z)\, \mathrm{d}\mu = \mu_Z(A - x).$$

And so, we find the corresponding action for a function $f \in C_0(\mathbb{R}^d)$, namely

$$Pf(x) = \int_{\mathbb{R}^d} f(z)\, P(x, \mathrm{d}z) = \int_{\mathbb{R}^d} f(z)\, \mu_Z(\mathrm{d}z - x) = \int_{\mathbb{R}^d} f(x + z)\, \mu_Z(\mathrm{d}z).$$

Now, we want to make this a continuous time process. To do this, we follow the same procedure as in example 2. We make our particle transition with a rate κ and call the path traced out by this continuous process $(X_t)_{t\geq 0}$. Now, as this process is in continuous

time, it is not characterised by a single time step transition function, but rather by a generator. Then, as shown in the same example, the generator of the process becomes

$$Lf(x) = \kappa(P - I)f(x) = \kappa \int_{\mathbb{R}^d} [f(x + z) - f(x)] \, \mu_Z(\mathrm{d}z).$$

Notice that in the discrete case we have $X_n - X_0 = \sum_{i=1}^{n} Z_i$, where $(Z_i)_{i \in \mathbb{N}}$ are i.i.d. copies of Z. In the continuous time case, we can count the number of jumps that have occurred using a rate κ Poisson process N_t and we find that $X_t - X_0 = \sum_{i=1}^{N_t} Z_i$.

3.2 Brownian Motion

Before taking a look at the scaling limit of a random walk, we must first specify what Brownian motion is exactly.

Definition 3.2. *Let $B = (B_t)_{t \geq 0}$ be a continuous time stochastic process and D a symmetric, positive semi-definite matrix. Then, B is called Brownian motion with diffusion matrix D if*

(i) $B_0 = 0$;

(ii) *for all $s < t$, $B_t - B_s \sim \mathcal{N}(0, (t - s)D)$ independently of all B_u with $u \leq s$;*

(iii) *$t \mapsto B_t(\omega)$ is continuous for almost all $\omega \in \Omega$;*

where $\mathcal{N}(0, (t - s)D)$ denotes a multivariate normal distribution with covariance matrix $(t - s)D$.

So, Brownian motion is a stochastic process which start at 0, has normally distributed increments, which are independent if they do not overlap, and for each realisation the path $B_t(\omega)$ is continuous as a function of time. Moreover, if $D = I$, then we speak of standard Brownian motion.

Note that by the second property of the above definition, B is a Markov process. We then might wonder what the generator of Brownian motion is. To do this, suppose we start at some position x, so $B_0 = x$. Then, since $B_t = B_t - B_0 \sim \mathcal{N}(0, (t-0)D)$, we have

$$S_t(x, A) = \mu(x + B_t \in A) = \mu(x + \mathcal{N}(0, tD) \in A).$$

And so our semigroup action for $f \in C_0^\infty(\mathbb{R}^d)$ is given by

$$S_t f(x) = \mathbb{E}[f(x + B_t)] = \mathbb{E}[f(x + \mathcal{N}(0, tD))].$$

Now, if we apply a Taylor expansion to f and note that $\mathbb{E}[o(\mathcal{N}(0, tD)^2)] = o(t)$, we find in Einstein summation notation

$$\begin{aligned} S_t f(x) &= \mathbb{E}\left[f(x) + \mathcal{N}(0, tD)^i \partial_i f(x) + \frac{1}{2} \mathcal{N}(0, tD)^i \mathcal{N}(0, tD)^j \partial_i \partial_j f(x) \right] + o(t) \\ &= f(x) + \frac{1}{2} \mathrm{Cov}(\mathcal{N}(0, tD))^{ij} \partial_i \partial_j f(x) + o(t) \\ &= f(x) + \frac{1}{2} t D^{ij} \partial_i \partial_j f(x) + o(t), \end{aligned}$$

where it is used that $\mathbb{E}[\mathcal{N}(0,tD)] = 0$ and $\mathbb{E}[\mathcal{N}(0,tD)\mathcal{N}(0,tD)^T] = \mathrm{Cov}(\mathcal{N}(0,tD)) = tD$. Using the above, we find that the generator is given by

$$\begin{aligned} Lf(x) &= \lim_{t\to 0} \frac{S_t - I}{t} f(x) \\ &= \lim_{t\to 0} \frac{\frac{1}{2} tD^{ij}\partial_i\partial_j f(x) + o(t)}{t} \\ &= \frac{1}{2} D^{ij}\partial_i\partial_j f(x). \end{aligned}$$

Note, in particular, that if $D = I$ that the generator is the the Laplace operator; $L = \nabla^2$. This establishes a link between the diffusion equation and Brownian motion, since $S_t f(x)$ is the unique solution to the differential equation $\frac{\mathrm{d}}{\mathrm{d}t}\psi_t = Lg_t = \nabla^2\psi_t$ with $\psi_0 = f$ by theorem 2.8 (ii). This is precisely the diffusion equation and thus solving the diffusion equation corresponds to evolving the initial condition by the semigroup. In chapter 6, we will use this relation in the development of Diffusion Monte Carlo.

Now, as a last point of interest, we note that we could also add a drift to our normal distribution, so that we find $B_t - B_s \sim \mathcal{N}((t-s)m, (t-s)D)$, where m is the drift velocity. As drift is a phenomenon that we will encounter multiple times later on, we will simply find the generator for a particle experiencing an independent drift. Assume we have a particle X_t, which experiences a constant drift m. Then, we find using a Taylor expansion that

$$\begin{aligned} Lf(x) &= \lim_{t\to 0} \frac{S_t f(x) - f(x)}{t} \\ &= \lim_{t\to 0} \frac{f(x+mt) - f(x)}{t} \\ &= \lim_{t\to 0} m^i\partial_i f(x) + o(1) \\ &= m^i\partial_i f(x). \end{aligned}$$

Then, since for Brownian motion drift is indepedent of the diffusion, we find the generator

$$Lf(x) = m^i\partial_i f(x) + \frac{1}{2} D^{ij}\partial_i\partial_j f(x).$$

3.3 Scaling

The main type of scaling that is of interest in this thesis is diffusive scaling. This scaling closely resembles the central limit theorem and in particular we can show that under this scaling a random walk converges to Brownian motion.

Definition 3.3. *Let $(X_t)_{t\geq 0}$ be an $\mathbb{R}^d$-valued Feller process. Then the diffusive scaling of this process is defined as $\lim_{\varepsilon\to 0} \varepsilon X_{\varepsilon^{-2}t}$ if it exists. Here, this limit is in the sense of weak convergence in the path space.*

Thus under diffusive scaling the position of the particle is scaled by ε i.e. $x \mapsto \varepsilon x$ and the time of the process is scaled by ε^{-2} i.e. $t \mapsto \varepsilon^{-2}t$. Thus on the one hand, our particles trajectory shrinks, but on the other hand, it travels through its trajectory faster.

Remark 1. As we have already mentioned, Brownian motion is special when it comes to diffusive scaling, since it is invariant under this scaling. So for a Brownian motion process B_t, $\varepsilon B_{\varepsilon^{-2}t} \to B_t$ as $\varepsilon \to 0$, in fact $\varepsilon B_{\varepsilon^{-2}t} \overset{d}{=} B_t$ for all $\varepsilon > 0$.

Furthermore, diffusive scaling also allows us to analyse the long term behaviour of our process. If we zoom out, we expect it to behave similar to the Brownian motion process it scales to. This is quite similar what to what the central limit theorem does, which states that a rescaled sum of i.i.d. L_2 random variables converges to a normal distribution. $\triangle$

Now, we want to show that the random walk from section 3.1 converges to Brownian motion with respect to diffusive scaling.

Proposition 3.4. *Let $(X_t)_{t\geq 0}$ be a random walk jumping along the $\mathbb{R}^d$-valued random variable $Z \in L^2$ at rate κ. Moreover, assume $\mathbb{E}[Z] = 0$. Then, $\varepsilon X_{\varepsilon^{-2}t}$ converges to Brownian motion B_t with diffusion matrix $\kappa Cov(Z)$.*

Proof. By theorem 2.10, it suffices to show that the generator of $\varepsilon X_{\varepsilon^{-2}t}$ converges to a generator of Brownian motion. Thus, we first need to find the generator L_ε of $\varepsilon X_{\varepsilon^{-2}t}$.

Note that the scaling of the position of the particle essentially corresponds to scaling the jumps by ε. Thus we find that $x + Z \mapsto x + \varepsilon Z$ and so $P_\varepsilon f(x) = \int_{\mathbb{R}^d} f(x + \varepsilon z)\, \mu_Z(\mathrm{d}z)$. Let us now denote the Poisson process counting the cumulative number of jumps by N_t. Then in the same fashion as in example 2 only now scaling t by ε^{-2}, the semigroup for the scaled process becomes

$$\begin{aligned} S_{\varepsilon^{-2}t} f(x) &= \sum_{n=0}^{\infty} \mathbb{P}(N_{\varepsilon^{-2}t} = n) P_\varepsilon^n f(x) \\ &= f(x)(1 - \kappa\varepsilon^{-2}t) + P_\varepsilon f(x)\kappa\varepsilon^{-2}t + o(t). \end{aligned}$$

Thus, the generator is of the scaled process is given by

$$\begin{aligned} L_\varepsilon f(x) &= \lim_{t\to 0} \frac{S_{\varepsilon^{-2}t} f(x) - f(x)}{t} \\ &= \lim_{t\to 0} \varepsilon^{-2}\kappa(P_\varepsilon - I) f(x) + o(1) \\ &= \varepsilon^{-2}\kappa(P_\varepsilon - I) f(x). \end{aligned} \tag{1}$$

So, for the scaled random walk, we find $L_\varepsilon f(x) = \varepsilon^{-2}\kappa \int_{\mathbb{R}^d} [f(x + \varepsilon z) - f(x)]\, \mu_Z(\mathrm{d}z)$.

To show that the scaled random walk converges to Brownian motion, we choose $f \in C_0^\infty$ and compute the limit $\lim_{\varepsilon\to 0} L_\varepsilon f$. It then suffices for this limit to coincide with a generator of Brownian motion by theorem 2.10. Now, by applying a Taylor expansion to f, we find using $\mathbb{E}[Z] = 0$ and $\mathbb{E}[ZZ^T] = \mathrm{Cov}(Z)$ that

$$\begin{aligned} L_\varepsilon f(x) &= \varepsilon^{-2}\kappa \int_{\mathbb{R}^d} [\varepsilon z^i \partial_i f(x) + \frac{1}{2}\varepsilon^2 z^i z^j \partial_i\partial_j f(x) + o(\varepsilon^2)]\, \mu_Z(\mathrm{d}z) \\ &= \varepsilon^{-1}\kappa \mathbb{E}[Z]^i \partial_i f(x) + \frac{1}{2}\kappa \mathrm{Cov}(Z)^{ij}\partial_i\partial_j f(x) + o(1) \\ &= \frac{1}{2}\kappa \mathrm{Cov}(Z)^{ij}\partial_i\partial_j f(x) + o(1) \\ &\longrightarrow \frac{1}{2}\kappa \mathrm{Cov}(Z)^{ij}\partial_i\partial_j f(x) \text{ as } \varepsilon \to 0. \end{aligned}$$

Thus, a random walk which jumps according to the random variable Z, converges in the diffusive limit to a Brownian motion with diffusion matrix $\kappa \mathrm{Cov}(Z)$. □

Apart from the definition of Brownian motion we have given, there exist various other ways of characterising Brownian motion. One such way is through the Lévy characterisation, which is proven in [7].

Theorem 3.5. *A $\mathbb{R}^d$-valued process $B = (B_t)_{t \geq 0}$ with continuous sample paths corresponds to Brownian motion with diffusion matrix D if and only if for all $i, j \in \{1, ..., d\}$,*

(i) *B_t^i is a martingale for all $t \geq 0$;*

(ii) *$[B^i, B^j]_t = D^{ij} t$,*

where B^i denotes the ith component of the process.

4 Active particles in $\mathbb{R}^d$

In this chapter, we start with the core matter of this thesis. Firstly, an introduction to active particles is given, specifying what they actually entail. Thereafter, a proof is given that the active particle process converges to Brownian motion in the diffusive limit. This proof bares a strong resemblance with the proof from [1]. After that, a new proof of the same convergence result is introduced, which generalises more easily to manifolds. Finally, some examples of active particle motion are worked out.

4.1 Model

We will start of with a simple example of an active particle process. Then, we will formally specify what exactly the active particles are by defining their state space and dynamics. Thereafter, we use the resulting ideas to derive a generator for these particles. Moreover, for all of this and the following sections, it is assumed that there is some joint underlying probability space $(\Omega, \mathscr{A}, \mu)$.

Example 4. We will explore a simple active particle process in $\mathbb{Z}$. This process consists of a location process $(X_t)_{t\geq 0}$ in $\mathbb{Z}$ and an internal state process in $(\sigma_t)_{t\geq 0}$ in $\{-1, 1\}$. Firstly, the internal state switches between states at rate γ. Now, the location process is subject to two independent transitions. With a rate λ, the particle jumps in the direction of the internal state σ_t. Furthermore, with rate κ, we have a random walk with a particle that jumps 1 space to the left or the right with equal probability. The generator of this process is given by

$$\begin{aligned} Lf(x,\sigma) = \frac{\kappa}{2}&[f(x-1,\sigma) + f(x+1,\sigma) - 2f(x,\sigma)] \\ &+ \lambda[f(x+\sigma,\sigma) - f(x,\sigma)] \\ &+ \gamma[f(x,-\sigma) - f(x,\sigma)]. \end{aligned}$$

We can view this process as a random walk with a drift $+\lambda$, when $\sigma = +1$ and as a random walk with a drift $-\lambda$, when $\sigma = -1$. △

For the general active particle process, we consider the location $X = (X_t)_{t\geq 0}$ of a particle in $\mathbb{R}^d$ with an internal state process $\sigma = (\sigma_t)_{t\geq 0}$ in the state space Σ equipped with the σ-algebra $\mathscr{S}$. Thus, the active particle process corresponds to the process $(X_t, \sigma_t)_{t\geq 0}$ in $\mathbb{R}^d \times \Sigma$. Furthermore, we have a measurable map $v : \Sigma \to \mathbb{R}^d$, which we call the velocity map.

Now, the internal state process $(\sigma_t)_{t\geq 0}$ is assumed to be a Feller process with semigroup $(S_t)_{t\geq 0}$ and generator A. Now, we also assume this process starts from a unique stationary measure ν and is ergodic. From this, it follows that

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(\sigma_t)\,\mathrm{d}t \to \int_\Sigma f(\varsigma)\nu(\mathrm{d}\varsigma).$$

Furthermore, we assume that for all $f \in D(A)$, we have $\|S_t f - \int f(\varsigma)\,\nu(\mathrm{d}\varsigma)\|_{L^2(\nu)} \leq Ke^{-Ct}$, where $K, C \in \mathbb{R}$. Then, A satisfies the conditions from proposition 2.9 (iii), so A^{-1} exists for all $f \in D(A)$.

With the internal state process defined, we move on to the location process. This process starts at the origin, so $X_0 = 0$. Thereafter, the process evolves according to two independent jump processes, which are independent from the internal state process. Firstly, a random walk part at rate κ according to the jump vector Z, which has – as per section 3.1 – the generator

$$\kappa \int_{\mathbb{R}^d} [f(x+z,\sigma) - f(x,\sigma)]\, \mu_Z(\mathrm{d}z).$$

Secondly, we jump along the velocity vector $v(\sigma)$ at rate λ depending on the current internal state $\sigma \in \Sigma$. From this, it immediately follows that the generator is given by

$$\lambda[f(x+v(\sigma),\sigma) - f(x,\sigma)].$$

In addition, we will assume that v is $L^2(\nu)$ and $\mathbb{E}_\nu[v(\sigma)] = \int_\Sigma v(\varsigma)\nu(\mathrm{d}\varsigma) = 0$, which can be interpreted as the absent of drift. If we relax this condition, we will observe that the velocity map could induce a drift – an constant velocity component in the movement of the particle. Furthermore, this assumption ensures $v \in D(A)$ and thus, $A^{-1}v$ exists.

Thus to summarise, we have the following three independent jump processes:

(i) A random walk jump at rate κ

(ii) An active motion jump along the velocity vector $v(\sigma_t)$ at rate λ

(iii) An internal state transition according to its own independent Markov process with generator A

So, an active particle process is a random walk with an additional drift component, which is dictated by the internal state of the particle. This actually has a physical interpretation. On a microscopic scale, we can say that the random walk part can be associated with random collision between the active particle and neighbouring molecules and that the active motion is the result of some internal chemical state, which the active particle has.

Now, the generator L for the complete active particle process is quite simply the sum of the generators of the individual processes as per example 2, since the above three processes are independent. Furthermore, we will make a slight modification to the internal state process by adding a rate γ. Thus, we take $(\sigma_{\gamma t})_{t\geq 0}$ as the internal state process and from now on out just call it σ_t again, in the knowledge that the process depends on γ. This rate can be used to adjust the speed of the internal process allowing us to take the relative speed of the internal process into account. Then, it follows that the generator of the active particle process is the following:

$$\begin{aligned} Lf(x,\sigma) = \kappa &\int_{\mathbb{R}^d} [f(x+z,\sigma) - f(x,\sigma)]\, \mu_Z(\mathrm{d}z) \\ &+ \lambda[f(x+v(\sigma),\sigma) - f(x,\sigma)] \\ &+ \gamma A f(x,\sigma). \end{aligned} \tag{2}$$

Later on, we will also need the generator L_ε of the diffusively scaled process $(\varepsilon X_{\varepsilon^{-2}t}, \sigma_{\varepsilon^{-2}t})_{t\geq 0}$, which follows analogously to equation 1 and is given by

$$\begin{aligned} L_\varepsilon f(x,\sigma) = \varepsilon^{-2}\kappa &\int_{\mathbb{R}^d} [f(x+\varepsilon z,\sigma) - f(x,\sigma)]\, \mu_Z(\mathrm{d}z) \\ &+ \varepsilon^{-2}\lambda[f(x+\varepsilon v(\sigma),\sigma) - f(x,\sigma)] \\ &+ \varepsilon^{-2}\gamma A f(x,\sigma). \end{aligned} \tag{3}$$

4.2 Martingale Scaling

Having established the active particle process X_t in the previous section, we would now like to find the diffusive limit of this process $\varepsilon X_{\varepsilon^{-2}t}$. One way to find this limit is by using martingales.

Theorem 4.1. *Let $(X_t, \sigma_t)_{t\geq 0}$ be an active particle process in $\mathbb{R}^d$ with a generator as in equation 2. Then the diffusive limit of X converges to Brownian motion with diffusion matrix $D^{ij} = \kappa Cov(Z)^{ij} + \lambda Cov(v)^{ij} + \frac{\lambda^2}{\gamma}\int_\Sigma v^i_\varsigma(-A^{-1})(v^j_\varsigma) + v^j_\varsigma(-A^{-1})(v^i_\varsigma)\nu(\mathrm{d}\varsigma)$.*

The following proof is based on the proof given in [1].

Proof. We will, firstly, show that the diffusive limit of $(X_t, \sigma_t)_{t\geq 0}$ is Brownian motion. Thereafter, we will compute the diffusion matrix. To show the convergence to Brownian motion, we will write X_t as a martingale and something negligible. Then, we can use a theorem on the diffusive limit of martingales with stationary ergodic increments to show that the diffusive limit of X_t is Brownian motion.

In what follows, we will use i, j to denote arbitrary indices from $\{1, ..., d\}$. We start out by using $\pi_i : \mathbb{R}^d \times \Sigma \to \mathbb{R}$ to denote the projection of the ith component. Then, we know from theorem 2.14, that M^{π_i} is a martingale. Now, firstly note that

$$\begin{aligned}
L_\varepsilon(\pi_i) &= \varepsilon^{-2}\kappa\int_{\mathbb{R}^d}\pi_i(x+\varepsilon z,\sigma) - \pi_i(x,\sigma)\mu_Z(\mathrm{d}z)\\
&\quad + \varepsilon^{-2}\lambda(\pi_i(x+\varepsilon v(\sigma),\sigma) - \pi_i(x,\sigma))\\
&\quad + \varepsilon^{-2}\gamma A\pi_i(x,\sigma)\\
&= \varepsilon^{-2}\left[\kappa\int_{\mathbb{R}^d}\varepsilon z^i\mu_Z(\mathrm{d}z) + \lambda\varepsilon v^i(\sigma) + \gamma A x^i\right]\\
&= \varepsilon^{-1}\lambda v^i(\sigma),
\end{aligned}$$

where in the final equality, it has been used that $E[Z] = 0$ and A only acts on σ, so per proposition 2.9, $Ax^i = 0$. Then, it follows that

$$\begin{aligned}
M^{\pi_i}_{\varepsilon,t} &= \pi_i(\varepsilon X_{\varepsilon^{-2}t}) - \pi_i(\varepsilon X_0) - \int_0^t L_\varepsilon(\pi_i)(\varepsilon X_{\varepsilon^{-2}s}, \sigma_{\varepsilon^{-2}s})\,\mathrm{d}s\\
&= \varepsilon X^i_{\varepsilon^{-2}t} - 0 - \varepsilon^{-1}\lambda\int_0^t v^i(\sigma_{\varepsilon^{-2}s})\,\mathrm{d}s\\
&= \varepsilon X^i_{\varepsilon^{-2}t} - \varepsilon\lambda\int_0^{\varepsilon^{-2}t} v^i(\sigma_u)\,\mathrm{d}u\,,
\end{aligned}$$

where the substitution $u = \varepsilon^{-2}s$ has been used and the fact that $X_0 = 0$. So, we see that

$$\varepsilon X^i_{\varepsilon^{-2}t} = M^{\pi_i}_{\varepsilon,t} + \varepsilon\lambda\int_0^{\varepsilon^{-2}t} v^i(\sigma_u)\,\mathrm{d}u\,.$$

Now, since $\sigma_{\varepsilon^{-2}t}$ is a Markov process, we can apply theorem 2.14 once again, this time with $g_i(\sigma) = \epsilon\lambda A^{-1}v^i(\sigma)$, where we recall that $A^{-1}v(\sigma)$ exists since $\mathbb{E}_\nu[v(\sigma)] = 0$ and

σ_t goes quickly to its stationary distribution. So then we find that the following is a martingale:

$$\begin{aligned} M^{g_i}_{\varepsilon,t} &= \varepsilon\lambda A^{-1}v^i(\sigma_{\varepsilon^{-2}s}) - \varepsilon\lambda A^{-1}v^i(\sigma_0) - \varepsilon\lambda\int_0^t AA^{-1}v^i(\sigma_{\varepsilon^{-2}s})\,\mathrm{d}s \\ &= \varepsilon\lambda A^{-1}(v^i(\sigma_{\varepsilon^{-2}t}) - v^i(\sigma_0)) - \varepsilon\lambda\int_0^{\varepsilon^{-2}t} v^i(\sigma_u)\,\mathrm{d}u\,. \end{aligned}$$

And so, we have that

$$\varepsilon\lambda\int_0^{\varepsilon^{-2}t} v^i(\sigma_u)\,\mathrm{d}u = -M^{g_i}_{\varepsilon,t} + \varepsilon\lambda A^{-1}(v^i(\sigma_{\varepsilon^{-2}t}) - v^i(\sigma_0))$$

Thus, we find that

$$\varepsilon X^i_{\varepsilon^{-2}t} = M^{\pi_i}_{\varepsilon,t} - M^{g_i}_{\varepsilon,t} + \varepsilon\lambda A^{-1}(v^i(\sigma_{\varepsilon^{-2}t}) - v^i(\sigma_0)). \tag{4}$$

By assumption, we have that the internal state process is in the stationary distribution ν, so $\sigma_{\varepsilon^{-2}t} \sim \nu$ and $\|A^{-1}v^i(\sigma_{\varepsilon^{-2}t})\|^2_{L^2(\nu)} = \int_\Sigma (A^{-1}v^i(\varsigma))^2\,\nu(\mathrm{d}\varsigma) < \infty$. Thus, we have that the term $\varepsilon\lambda A^{-1}(v^i(\sigma_{\varepsilon^{-2}t}) - v^i(\sigma_0))$ from equation 4 vanishes in $L^2(\nu)$ as $\varepsilon \to 0$.

Then, if we let M^i_t and $\tilde{M}^i_t$ be the diffusive limits of the martingales $M^{\pi_i}_{\varepsilon,t}$ and $M^{g_i}_{\varepsilon,t}$ respectively, we find that as $\varepsilon \to 0$,

$$\varepsilon X^i_{\varepsilon^{-2}t} \to Z_t = M^i_t + \tilde{M}^i_t.$$

The diffusive limit of these martingales converges to Brownian motion by [11]. Moreover, as shown in [1], the covariance of the above two martingales is 0. Then, it follows that Z_t is once again Brownian motion and the diffusion matrix of Z_t is the sum of the diffusion matrices of M_t and $\tilde{M}_t$.

So, to compute the diffusion matrix of Z_t, we just have to compute the diffusion of M_t and $\tilde{M}_t$. Firstly, observe that

$$\begin{aligned} L_\varepsilon(\pi_i\pi_j) &= \varepsilon^{-2}\kappa\int_{\mathbb{R}^d} \pi_i(x+\varepsilon z,\sigma)\pi_j(x+\varepsilon z,\sigma) - \pi_i(x,\sigma)\pi_j(x,\sigma)\mu_Z(\mathrm{d}z) \\ &\quad + \varepsilon^{-2}\lambda(\pi_i(x+\varepsilon v(\sigma),\sigma)\pi_j(x+\varepsilon v(\sigma),\sigma) - \pi_i(x,\sigma)\pi_j(x,\sigma)) \\ &\quad + \varepsilon^{-2}\gamma A\pi_i(x,\sigma)\pi_j(x,\sigma) \\ &= \varepsilon^{-2}\kappa\int_{\mathbb{R}^d} \varepsilon z(x^i+x^j) + \varepsilon^2 z^i z^j\mu_Z(\mathrm{d}z) + \varepsilon^{-2}\lambda(\varepsilon x^i v^j(\sigma) + \varepsilon x^j v^i(\sigma) + \varepsilon^2 v^i v^j) \\ &\quad + \varepsilon^{-2}\gamma A x^i x^j \\ &= \kappa\int_{\mathbb{R}^d} z^i z^j\mu_Z(\mathrm{d}z) + \lambda v^i v^j + \varepsilon^{-1}\lambda(x^i v^j(\sigma) + x^j v^i(\sigma)) \\ &= \kappa\mathrm{Cov}(Z)^{ij} + \lambda v^i v^j + \varepsilon^{-1}\lambda(x^i v^j(\sigma) + x^j v^i(\sigma)), \end{aligned}$$

were we have used that $\mathbb{E}[Z] = 0$ and A only acts on σ. Then using theorem 2.15 and

recalling that $\pi_i L_\varepsilon \pi_j = \varepsilon^{-1} x^i \lambda v^i(\sigma)$, we find that

$$\begin{aligned}
[M_\varepsilon^{\pi_i}, M_\varepsilon^{\pi_j}]_t &= \int_0^t \left[L_\varepsilon(\pi_i\pi_j) - \pi_i L_\varepsilon \pi_j - \pi_j L_\varepsilon \pi_i\right](\varepsilon X_{\varepsilon^{-2}s}, \sigma_{\varepsilon^{-2}s})\,\mathrm{d}s \\
&= \int_0^t [\kappa Cov(Z)^{ij} + \lambda v^i v^j + \varepsilon^{-1}\lambda(x^i v^j(\sigma) + x^j v^i(\sigma)) \\
&\quad - \varepsilon^{-1} x^i \lambda v^j(\sigma) - \varepsilon^{-1} x^j \lambda v^i(\sigma)](\varepsilon X_{\varepsilon^{-2}s}, \sigma_{\varepsilon^{-2}s})\,\mathrm{d}s \\
&= k\mathrm{Cov}(Z)^{ij} t + \varepsilon^2 \lambda \int_0^{\varepsilon^{-2}t} (v^i v^j)(\sigma_u)\,\mathrm{d}u\,,
\end{aligned}$$

where the substitution $u = \varepsilon^{-2}s$ was used. So, if we now let $\varepsilon \to 0$, we know using the ergodicity of σ_t that

$$\varepsilon^2 \int_0^{\varepsilon^{-2}t} (v^i v^j)(\sigma_u)\,\mathrm{d}u \to t \int_\Sigma v^i(\varsigma) v^j(\varsigma)\,\nu(\mathrm{d}\varsigma).$$

So, we find that

$$\begin{aligned}
[M, M]_t &= k\mathrm{Cov}(Z)^{ij} t + t\lambda \int_\Sigma v^i(\varsigma) v^j(\varsigma)\,\nu(\mathrm{d}\varsigma) \\
&= [k\mathrm{Cov}(Z)^{ij} + \lambda \mathrm{Cov}(v)^{ij}]t.
\end{aligned}$$

Thus, by the lévy characterisation, the diffusion matrix associated with M_t is $k\mathrm{Cov}(Z)^{ij} + \lambda\mathrm{Cov}(v)^{ij}$.

Now, we also want to calculate the diffusion matrix associated with $\tilde{M}_t$. As shown before $\tilde{M}_t$ is Brownian motion, we will exploit this knowledge here to compute the diffusion matrix of $\tilde{M}_t$. Since Brownian motion with diffusion matrix D at time t has distribution $\mathcal{N}(0, tD)$, we know that under diffusive scaling $C^{ij}(t) = \mathrm{Cov}(\lambda \int_0^t v^i(\sigma_s)\,\mathrm{d}s\,, \lambda \int_0^t v^j(\sigma_r)\,\mathrm{d}r)$ converges to tD^{ij}. So, using time homogeneity, we see

$$\begin{aligned}
C^{ij}(t) &= \lambda^2 \mathbb{E}_\nu \left[\int_0^t v^i(\sigma_s)\,\mathrm{d}s \int_0^t v^j(\sigma_r)\,\mathrm{d}r\right] \\
&= \lambda^2 \int_0^t \int_0^t \mathbb{E}_\nu[v^i(\sigma_s) v^j(\sigma_r)]\,\mathrm{d}s\,\mathrm{d}r \\
&= \lambda^2 \int_0^t \left[\int_0^r \mathbb{E}_\nu[v^i(\sigma_s) v^j(\sigma_r)]\,\mathrm{d}s + \int_r^t \mathbb{E}_\nu[v^i(\sigma_s) v^j(\sigma_r)]\,\mathrm{d}s\right]\mathrm{d}r \\
&= \lambda^2 \int_0^t \left[\int_0^r \mathbb{E}_\nu[v^i(\sigma_0) v^j(\sigma_{r-s})]\,\mathrm{d}s + \int_r^t \mathbb{E}_\nu[v^i(\sigma_{s-r}) v^j(\sigma_0)]\,\mathrm{d}s\right]\mathrm{d}r \\
&= \lambda^2 \int_0^t \left[-\int_r^0 \mathbb{E}_\nu[v^i(\sigma_0) v^j(\sigma_u)]\,\mathrm{d}u + \int_0^{t-r} \mathbb{E}_\nu[v^i(\sigma_u) v^j(\sigma_0)]\,\mathrm{d}u\right]\mathrm{d}r \\
&= \lambda^2 \int_0^t \int_0^r \mathbb{E}_\nu[v^i(\sigma_0) v^j(\sigma_u)]\,\mathrm{d}u\,\mathrm{d}r - \int_t^0 \int_0^r \mathbb{E}_\nu[v^i(\sigma_u) v^j(\sigma_0)]\,\mathrm{d}u\,\mathrm{d}r \\
&= \lambda^2 \int_0^t \int_0^r \mathbb{E}_\nu[v^i(\sigma_0) v^j(\sigma_u) + v^i(\sigma_u) v^j(\sigma_0)]\,\mathrm{d}u\,\mathrm{d}r\,.
\end{aligned}$$

Applying diffusive scaling to the above equation, we find that

$$\begin{aligned}
C_\varepsilon^{ij}(t) &= \lambda^2\varepsilon^2 \int_0^{\varepsilon^{-2}t} \int_0^r \mathbb{E}_\nu[v^i(\sigma_0)v^j(\sigma_u) + v^i(\sigma_u)v^j(\sigma_0)]\,\mathrm{d}u\,\mathrm{d}r \\
&= \lambda^2\varepsilon^2 \int_0^{\varepsilon^{-2}t} \int_u^{\varepsilon^{-2}t} \mathbb{E}_\nu[v^i(\sigma_0)v^j(\sigma_u) + v^i(\sigma_u)v^j(\sigma_0)]\,\mathrm{d}r\,\mathrm{d}u \\
&= \lambda^2\varepsilon^2 \int_0^{\varepsilon^{-2}t} \mathbb{E}_\nu[v^i(\sigma_0)v^j(\sigma_u) + v^i(\sigma_u)v^j(\sigma_0)] \int_u^{\varepsilon^{-2}t} \mathrm{d}r\,\mathrm{d}u \\
&= \lambda^2 \int_0^\infty \mathbb{E}_\nu[v^i(\sigma_0)v^j(\sigma_u) + v^i(\sigma_u)v^j(\sigma_0)]\mathbf{1}_{[0,\varepsilon^{-2}t]}(u)(t-\varepsilon^2 u)\,\mathrm{d}u\,,
\end{aligned}$$

where $\mathbf{1}_A$ denotes the indicator function of the set A. Note that $|\mathbb{E}_\nu[v^i(\sigma_0)v^j(\sigma_u) + v^i(\sigma_u)v^j(\sigma_0)]\mathbf{1}_{[0,\varepsilon^{-2}t]}(u)|t \le |\mathbb{E}_\nu[v^i(\sigma_0)v^j(\sigma_u) + v^i(\sigma_u)v^j(\sigma_0)]|t$, which is integrable since $\|\int_0^\infty S_t v(\varsigma)\,\nu(\mathrm{d}\varsigma)\|_{L^2(\nu)} \le \int_0^\infty \|S_t v(\varsigma)\|_{L^2(\nu)}\,\nu(\mathrm{d}\varsigma) \le K\int_0^\infty e^{-Ct}\,\mathrm{d}t < \infty$. Furthermore, note that $|\mathbf{1}_{[0,\varepsilon^{-2}t]}(u)\varepsilon^2 u| \le t$. Thus, $|\mathbb{E}_\nu[v^i(\sigma_0)v^j(\sigma_u) + v^i(\sigma_u)v^j(\sigma_0)]\mathbf{1}_{[0,\varepsilon^{-2}t]}(u)\varepsilon^2 u| \le |\mathbb{E}_\nu[v^i(\sigma_0)v^j(\sigma_u) + v^i(\sigma_u)v^j(\sigma_0)]|t$, which is integrable. Thus, we can apply the dominated convergence theorem to find that as $\varepsilon \to 0$,

$$C_\varepsilon^{ij}(t) \to \lambda^2 \left[\int_0^\infty \mathbb{E}_\nu[v^i(\sigma_0)v^j(\sigma_u) + v^i(\sigma_u)v^j(\sigma_0)]\,\mathrm{d}u\right] t.$$

Then, applying Fubini's theorem and further simplifying the expressions using the definition of the inverse generator A^{-1}, we see that

$$\begin{aligned}
C_\varepsilon^{ij}(t) &\to \lambda^2 \left(\mathbb{E}_\nu\left[v^i(\sigma_0)\int_0^\infty v^j(\sigma_u)\,\mathrm{d}u\right] + \mathbb{E}_\nu\left[v^j(\sigma_0)\int_0^\infty v^i(\sigma_u)\,\mathrm{d}u\right]\right) t \\
&= \lambda^2 \left(\mathbb{E}_\nu\left[v^i(\sigma)(-\gamma A)^{-1}v^j(\sigma)\right] + \mathbb{E}_\nu\left[v^j(\sigma)(-\gamma A)^{-1}v^i(\sigma)\right]\right) t \\
&= \frac{\lambda^2}{\gamma}\left(\mathbb{E}_\nu\left[v^i(\sigma)(-A^{-1})v^j(\sigma)\right] + \mathbb{E}_\nu\left[v^j(\sigma)(-A^{-1})v^i(\sigma)\right]\right) t \\
&= \frac{\lambda^2}{\gamma}\int_\Sigma \left[v^i(\varsigma)(-A^{-1})v^j(\varsigma) + v^j(\varsigma)(-A^{-1})v^i(\varsigma)\right]\,\nu(\mathrm{d}\varsigma)\,t.
\end{aligned}$$

Thus, we find that $\tilde{M}_t$ has the diffusion matrix $\frac{\lambda^2}{\gamma}\int_\Sigma [v_\varsigma^i(-A^{-1})(v_\varsigma^j) + v_\varsigma^j(-A^{-1})(v_\varsigma^i)]\,\nu(\mathrm{d}\varsigma)$.

All in all, we see that $\varepsilon X_{\varepsilon^{-2}t}$ converges to Brownian motion with diffusion matrix $D^{ij} = \kappa\mathrm{Cov}(Z)^{ij} + \lambda\mathrm{Cov}(v)^{ij} + \frac{\lambda^2}{\gamma}\int_\Sigma [v_\varsigma^i(-A^{-1})(v_\varsigma^j) + v_\varsigma^j(-A^{-1})(v_\varsigma^i)]\,\nu(\mathrm{d}\varsigma)$. □

Thus, we see that the active particle process leads to a modified diffusion coefficient. Firstly, we have the usual random walk part; $\kappa\mathrm{Cov}(Z)$. However, in addition we have diffusion due to the variance of the velocity field in the stationary distribution ν, namely $\lambda\mathrm{Cov}(v)$.

Furthermore, we have the term $\lambda^2\int_\Sigma v_\varsigma^i(-A^{-1})(v_\varsigma^j) + v_\varsigma^j(-A^{-1})(v_\varsigma^i)\nu(\mathrm{d}\varsigma)$. This term is slightly more difficult to interpret, but it is the variance introduced by the autocovariance of the active motion. So the diffusion matrix is not solely dictated by the stationary distribution of the internal state, but the actual process associated with A also has some influence.

We can observe that as $\gamma \to \infty$, this term vanishes. This vanishing is caused by the fact that if the internal state process runs very quickly, the covariance between states

separated in time quickly vanishes. However, if $\gamma \to 0$, then this term blows up. Since the internal state process runs relatively slowly, it will take a long time for the process to become uncorrelated from its past states. So then, the variance introduced by this term will be large.

4.3 Homogenisation

The proof of theorem 4.1 given in the previous section relies on the Lévy characterisation. Now, if want to extend this proof to manifolds, we cannot directly use the Lévy characterisation, since we are using the vector space structure of $\mathbb{R}^d$, which a manifold does not have. And although there does exists an analogue of the Lévy characterisation and quadratic variation on manifolds for semimartingales, these notions are not very workable. Therefore, we will introduce a novel homogenisation technique for proving the convergence of the diffusive limit of active particle motion to Brownian motion through theorem 2.10. This technique turns out to be powerful enough to extend to the manifold setting.

Recall that in section 3.3, we showed that the diffusive limit of a random walk is Brownian motion by using generator convergence. In this case, we could simply consider the generator of the process L_ε applied to an arbitrary f and see that $L_\varepsilon f$ goes to a generator of Brownian motion. However, for active particles, this is not the case.

Let us take $d = 1$ for now. So, we consider an active particle in $\mathbb{R}$. Then in particular, we see that Taylor expanding the term $\varepsilon^{-2}\lambda(f(x + \varepsilon v(\sigma), \sigma) - f(x, \sigma))$ yields $\varepsilon^{-1}\lambda v(\sigma) f'(x) + \mathcal{O}(1)$. Note that this term blows up, so the limit $\varepsilon \to 0$ does not exist in this case. However, looking at theorem 2.10, we have an additional degree of freedom, namely we can choose a sequence $f_\varepsilon \to f$ and then it suffices to show that $L_\varepsilon f_\varepsilon$ converges to a generator of Brownian motion $\frac{1}{2}Df''$ to show that the diffusive limit of our process is a form of Brownian motion.

This brings us to what is meant by homogenisation. By homogenisation, we mean constructing a sequence of functions f_ε which converges to f and for which furthermore, $L_\varepsilon f_\varepsilon$ does converge. Thus, homogenisation is a technique for removing singularities. So in general, we will fix a function $f \in D(\frac{\mathrm{d}^2}{\mathrm{d}x^2})$, the domain of Brownian motion. Note that in particular, we then have that f is independent of the internal state of the process and only depends on the location of the process. Thereafter, we construct a sequence of functions $f_\varepsilon \in D(L_\varepsilon)$ such that $f_\varepsilon \to f$ and $L_\varepsilon f_\varepsilon \to \frac{1}{2}Df''(x)$.

Before tackling the most general case, we start with two simpler examples.

Example 5 (Two-state system on $\mathbb{R}$)**.** Let us start out by considering the active particle process in $\mathbb{R} \times \{1, -1\}$ with generator $Lf(x, \sigma) = \lambda\sigma \frac{\mathrm{d}}{\mathrm{d}x} f(x, \sigma) + \gamma(f(x, -\sigma) - f(x, \sigma))$. This is a particle that moves in direction σ with velocity λ and switches between the states $\{1, -1\}$ at rate γ. Then the generator of the diffusively scaled process becomes

$$L_\varepsilon f(x, \sigma) = \varepsilon^{-1}\lambda\sigma \frac{\mathrm{d}}{\mathrm{d}x} f(x, \sigma) + \varepsilon^{-2}\gamma(f(x, -\sigma) - f(x, \sigma)).$$

Now, fix $f \in D(\frac{\mathrm{d}^2}{\mathrm{d}x^2})$ and let $f_\varepsilon(x, \sigma) = f(x) + \varepsilon g(x, \sigma)$, where g is to be determined. If

we fill this into the above equation, we find that

$$\begin{aligned} L_\varepsilon f_\varepsilon(x,\sigma) &= \varepsilon^{-1}\lambda\sigma\frac{\mathrm{d}}{\mathrm{d}x}f_\varepsilon(x,\sigma) + \varepsilon^{-2}\gamma(f_\varepsilon(x,-\sigma) - f_\varepsilon(x,\sigma)) \\ &= \varepsilon^{-1}\lambda\sigma\frac{\mathrm{d}}{\mathrm{d}x}f(x) + \lambda\sigma\frac{\mathrm{d}}{\mathrm{d}x}g(x,\sigma) + \varepsilon^{-1}\gamma(g(x,-\sigma) - g(x,\sigma)). \end{aligned}$$

In order for this to converge as $\varepsilon \to 0$, we find the following condition

$$\lambda\sigma f'(x) + \gamma(g(x,-\sigma) - g(x,\sigma)) = 0,$$

where $f'(x)$ denotes the derivative with respect to x. If we now associate the generator A with $g(x,-\sigma) - g(x,\sigma)$, then we see that $A\sigma = -2\sigma$, so $A^n\sigma = (-2)^n\sigma$. Thus, we observe that $-A^{-1}\sigma = \int_0^\infty S_t\sigma\,\mathrm{d}t = \int_0^\infty e^{tA}\sigma\,\mathrm{d}t = \int_0^\infty \sum_{n=0}^\infty \frac{t^nA^n}{n!}\sigma\,\mathrm{d}t = \int_0^\infty \sum_{n=0}^\infty \frac{(-2t)^n}{n!}\sigma\,\mathrm{d}t = \int_0^\infty e^{-2t}\,\mathrm{d}t\,\sigma = \frac{1}{2}\sigma$, we find that

$$g(x,\sigma) = \frac{\lambda}{\gamma}(-A^{-1})\sigma f'(x) = \frac{\lambda}{2\gamma}\sigma f'(x).$$

So, in that case, we find that since $\sigma \in \{-1,1\}$, $L_\varepsilon f_\varepsilon(x,\sigma) = \frac{1}{2}\frac{\lambda^2}{\gamma}\sigma^2 f''(x) = \frac{1}{2}\frac{\lambda^2}{\gamma}f''(x)$ and so it follows that

$$L_\varepsilon f_\varepsilon(x,\sigma) \to \frac{1}{2}\frac{\lambda^2}{\gamma}f''(x),$$

where $f''(x)$ denotes the second order derivative of f. And thus the active particle processes converges to Brownian motion with diffusion coefficient $\frac{\lambda^2}{\gamma}$. Importantly, note that this diffusion constant does not depend on the internal state of the process, which is a requirement for convergence as the limit $\frac{1}{2}Df''(x)$ does not depend on σ. △

Example 6 (General system on $\mathbb{R}$)**.** We will now consider the same process as in the previous example, only now with a general state space Σ and internal state process with generator A and stationary measure ν. Moreover, we take a velocity map $v : \Sigma \times \mathbb{R} \to \mathbb{R}$ such that for each $\sigma \in \Sigma$, the section $v_\sigma(\cdot) := v(\sigma,\cdot)$ is a smooth vector field and $\int_\Sigma v_\varsigma(x)\nu(\mathrm{d}\varsigma) = 0$ for all $x \in \mathbb{R}$. In this case, we take the generator $Lf(x,\sigma) = \lambda v_\sigma(x)\frac{\mathrm{d}}{\mathrm{d}x}f(x,\sigma) + \gamma Af(x,\sigma)$. We can view this as a particle moving along the vector field v_σ for some $\sigma \in \Sigma$ and jumps to a new state in Σ at a rate γ. Then, the generator of the diffusively scaled process is given by

$$L_\varepsilon f(x,\sigma) = \varepsilon^{-1}\lambda v_\sigma(x)\frac{\mathrm{d}}{\mathrm{d}x}f(x,\sigma) + \varepsilon^{-2}\gamma Af(x,\sigma).$$

Now, fix $f \in D(\frac{\mathrm{d}^2}{\mathrm{d}x^2})$, which does not depend on σ, and let $f_\varepsilon(x,\sigma) = f(x) + \varepsilon g(x,\sigma) + \varepsilon^2 h(x,\sigma)$, where g, h are to be determined. So we see that for a general internal state process, we need an additional term to homogenise our function f. If we fill this into the above equation, we find that

$$\begin{aligned} L_\varepsilon f_\varepsilon(x,\sigma) &= \varepsilon^{-1}\lambda v_\sigma(x)\frac{\mathrm{d}}{\mathrm{d}x}f_\varepsilon(x,\sigma) + \varepsilon^{-2}\gamma Af_\varepsilon(x,\sigma) \\ &= \varepsilon^{-1}\lambda v_\sigma(x)\frac{\mathrm{d}}{\mathrm{d}x}f(x) + \lambda v_\sigma(x)\frac{\mathrm{d}}{\mathrm{d}x}g(x,\sigma) + \varepsilon\lambda v_\sigma(x)\frac{\mathrm{d}}{\mathrm{d}x}h(x,\sigma) \\ &+ \varepsilon^{-2}\gamma Af(x) + \varepsilon^{-1}\gamma Ag(x,\sigma) + \gamma Ah(x,\sigma) \\ &= \varepsilon^{-1}[\lambda v_\sigma(x)f'(x) + \gamma Ag(x,\sigma)] + \lambda v_\sigma(x)\frac{\mathrm{d}}{\mathrm{d}x}g(x,\sigma) + \gamma Ah(x,\sigma) + o(1), \end{aligned}$$

where $o(1)$ represents all that terms that go to 0 as $\varepsilon \to 0$ and in the final equality, it was used that $f(x)$ is constant with respect to A. Since the velocity vector field is not constant, it might induce a drift. Thus, this time we might have convergence to Brownian motion with drift, which has a generator $cf'(x) + \frac{1}{2}Df''(x)$ with drift velocity $c \in \mathbb{R}$ and diffusion coefficient $D \geq 0$.

Now, in order for $L_\varepsilon f_\varepsilon$ not to blow up in the limit, we must require that terms associated with ε^{-1} vanish. Thus, we find that

$$\lambda v_\sigma(x) f'(x) + \gamma A g(x, \sigma) = 0.$$

Now, with the assumption that $A^{-1}v_\sigma$ exists, we find that

$$g(x, \sigma) = \frac{\lambda}{\gamma}[-A^{-1}v_\sigma(x)]f'(x).$$

From this, it follows that

$$\begin{aligned} L_\varepsilon f_\varepsilon(x, \sigma) &= \lambda v_\sigma(x) \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\lambda}{\gamma}[-A^{-1}v_\sigma(x)]f'(x)\right] + \gamma A h(x, \sigma) + o(1) \\ &= \lambda v_\sigma(x)\left[\frac{\mathrm{d}}{\mathrm{d}x}\big(-A^{-1}\big)v_\sigma(x)\right] f'(x) + \frac{\lambda^2}{\gamma} v_\sigma(x)(-A^{-1})v_\sigma(x) f''(x) \\ &\quad + \gamma A h(x, \sigma) + o(1). \end{aligned}$$

Clearly, the above still depends on σ. However, we want this limit to converge to $cf'(x) + \frac{1}{2}Df''(x)$, which does not depend on sigma. So this suggest that we should pick h such that

$$\gamma A h(x, \sigma) = \left(c - \lambda v_\sigma\left[\frac{\mathrm{d}}{\mathrm{d}x}\big(-A^{-1}\big)v_\sigma\right]\right) f'(x) + \frac{1}{2}\left(D - \frac{2\lambda^2}{\gamma} v_\sigma(-A^{-1})v_\sigma\right) f''(x).$$

However, for this h to exist, we need to have that $h \in D(A)$. So then we must also have that $\int_\Sigma Ah(x, \varsigma)\nu(\mathrm{d}\varsigma) = 0$, where ν is the stationary measure of our internal process. Then by taking $f(x) = x$, it follows that

$$c = \lambda \int_\Sigma v_\varsigma \left[\frac{\mathrm{d}}{\mathrm{d}x}\big(-A^{-1}\big)v_\varsigma\right] \nu(\mathrm{d}\varsigma).$$

And furthermore, by taking $f(x) = \frac{1}{2}x^2$, we find that

$$D = \frac{2\lambda^2}{\gamma} \int_\Sigma v_\varsigma(-A^{-1})v_\varsigma \nu(\mathrm{d}\varsigma).$$

So overall, we find that

$$L_\varepsilon f_\varepsilon(x, \sigma) \to cf'(x) + \frac{1}{2}Df''(x),$$

which is a limit independent of σ. So the diffusive limit of the original process is Brownian motion with drift c and diffusion D, where c, D are given above. Firstly, observe that c and D are x dependent. This, however, is not unusual for diffusion processes. In physical reality, location dependent diffusion occurs quite often for example when temperature is

not constant throughout a medium. In addition note that if $v(\sigma, \cdot)$ is a constant vector field, we actually have $c = 0$ – thus no drift – and D is independent of location. So that would conform to what was seen in the previous example.

Furthermore, as a more concrete example, consider the internal state space $\Sigma = \{-1, 1\}$ with an internal state process which switches between these two states at rate γ. Moreover, we have a location dependent smooth vector field $v(x)$ dictating $v_\sigma = \sigma v$. Recalling from the previous example $-A^{-1}\sigma = \frac{1}{2\sigma}$, we find that

$$c = \lambda \int_\Sigma v(x)\sigma \left[\frac{\mathrm{d}}{\mathrm{d}x}\big(-A^{-1}\big)v(x)\sigma \right] \nu(\mathrm{d}\varsigma) = \frac{1}{2}\lambda v(x)v'(x).$$

Now, let us fix the internal state process of the original active walk to be in the stationary measure. Then, even though in every point the movement of our particle is perfectly symmetric, we still on a global level see a drift induced by the gradient of the vector field v. We note that by simply checking the 4 possible combinations of v, v', we see that the particle will move towards locations, where it tends to move faster in the sense that $|v|$ is greater. △

Having gone through the above two examples, we will now consider the general active particle process as in section 4.1 with one major modification, namely the velocity function becomes a collection of smooth vector fields as in example 6. So we say $v : \Sigma \times \mathbb{R}^d \to \mathbb{R}^d$ and write v_σ for the section $v(\sigma, \cdot)$. We then assume that every v_σ is a smooth vector field.

Theorem 4.2. *Let $(X_t, \sigma_t)_{t\geq 0}$ be an active particle process with a collection of smooth vector fields as velocity function. Then, the diffusive limit of this process converges to the process with generator $c^j\partial_j f(x) + \frac{1}{2}D^{ij}\partial_i\partial_j f(x)$ on $C_0^2(\mathbb{R}^d)$, where*

$$c^j = \int_\Sigma v_\varsigma^i \partial_i(-A^{-1})(v_\varsigma^j)\, \nu(\mathrm{d}\varsigma)$$

and

$$D^{ij} = \kappa Cov(Z)^{ij} + \lambda \int_\Sigma v_\varsigma^i v_\varsigma^j \, \nu(\mathrm{d}\varsigma) + \frac{\lambda^2}{\gamma}\int_\Sigma v_\varsigma^i(-A^{-1})(v_\varsigma^j) + v_\varsigma^j(-A^{-1})(v_\varsigma^i)\, \nu(\mathrm{d}\varsigma),$$

are smoothly varying coefficients.

In case, the matrix D is constant and $c = 0$, we have the exact same result as in theorem 4.1. And the active particle process converges to Brownian motion. Otherwise, we still have that process converges to something akin to Brownian motion. However, the diffusion matrix of this Brownian motion varies smoothly as a function of position and a drift, which also varies smoothly a function of position, has been introduced.

Although this is not the exact problem we were looking at, it is a well studied problem [12, 13]. Indeed in many real world applications, diffusion coefficients depend on location, as the diffusive medium or temperature with the medium varies with position.

Proof. Assume we have an arbitrary function f in the domain of Brownian motion $C_0^2(\mathbb{R}^d)$, which thusly does not depend on σ. Then, we consider the sequence of functions $f_\varepsilon(x, \sigma) =$

$f(x)+\varepsilon g(x,\sigma)+\varepsilon^2 h(x,\sigma)$, where g, h are to be determined. Now, we once again consider the generator L_ε as in equation 3, which does not change when v_σ depends on location. Then applying L_ε to f_ε, Taylor expanding in the first component of $f_\varepsilon(x,\sigma)$ and leaving out the x dependence of v_σ for readability, we find

$$\begin{aligned}
L_\varepsilon f_\varepsilon(x,\sigma) &= \varepsilon^{-2}\kappa \int_{\mathbb{R}^d} f_\varepsilon(x+\varepsilon z,\sigma) - f_\varepsilon(x,\sigma)\mu_Z(\mathrm{d}z) \\
&+ \varepsilon^{-2}\lambda(f_\varepsilon(x+\varepsilon v_\sigma,\sigma) - f_\varepsilon(x,\sigma)) \\
&+ \varepsilon^{-2}\gamma A f_\varepsilon(x,\sigma) \\
&= \varepsilon^{-2}\kappa \int_{\mathbb{R}^d} \varepsilon z^i \partial_i f_\varepsilon(x,\sigma) + \frac{1}{2}\varepsilon^2 z^i z^j \partial_i\partial_j f_\varepsilon(x,\sigma) + o(\varepsilon^2)\mu_Z(\mathrm{d}z) \\
&+ \varepsilon^{-2}\lambda(\varepsilon v_\sigma^i \partial_i f_\varepsilon(x,\sigma) + \frac{1}{2}\varepsilon^2 v_\sigma^i v_\sigma^j \partial_i\partial_j f_\varepsilon(x,\sigma) + o(\varepsilon^2)) \\
&+ \varepsilon^{-2} A f_\varepsilon(x,\sigma) \\
&= \varepsilon^{-1}\kappa \int_{\mathbb{R}^d} z^i \mu_Z(\mathrm{d}z)\partial_i f_\varepsilon(x,\sigma) + \frac{1}{2}\kappa \int_{\mathbb{R}^d} z^i z^j \mu_Z(\mathrm{d}z)\partial_i\partial_j f_\varepsilon(x,\sigma) \\
&+ \varepsilon^{-1}\lambda v_\sigma^i \partial_i f_\varepsilon(x,\sigma) + \frac{1}{2}\lambda v_\sigma^i v_\sigma^j \partial_i\partial_j f_\varepsilon(x,\sigma) \\
&+ \varepsilon^{-2}\gamma A f_\varepsilon(x,\sigma) + o(1) \\
&= \frac{1}{2}\left[\kappa \mathrm{Cov}(Z)^{ij} + \lambda v_\sigma^i v_\sigma^j\right]\partial_i\partial_j f_\varepsilon(x,\sigma) + \varepsilon^{-1}\lambda v_\sigma^i \partial_i f_\varepsilon(x,\sigma) \\
&+ \varepsilon^{-2}\gamma A f_\varepsilon(x,\sigma) + o(1).
\end{aligned}$$

If we now substitute $f_\varepsilon(x,\sigma) = f(x) + \varepsilon g(x,\sigma) + \varepsilon^2 h(x,\sigma)$, recollect all vanishing terms in $o(1)$ and note that $Af(x) = 0$, we find that

$$\begin{aligned}
L_\varepsilon f_\varepsilon(x,\sigma) = \frac{1}{2}\left[\kappa \mathrm{Cov}(Z)^{ij} + \lambda v_\sigma^i v_\sigma^j\right]\partial_i\partial_j f(x) + \varepsilon^{-1}\lambda v_\sigma^i \partial_i f(x) \\
+ \lambda v_\sigma^i \partial_i g(x,\sigma) + \varepsilon^{-1}\gamma A g(x,\sigma) + \gamma A h(x,\sigma) + o(1).
\end{aligned}$$

Firstly, if the above expression is to converge to the generator of Brownian motion for all f as $\varepsilon \to 0$, we need that the terms associated with ε^{-1} vanish. Thus,

$$\lambda v_\sigma^i \partial_i f(x) + \gamma A g(x,\sigma) = 0 \implies g(x,\sigma) = -\frac{\lambda}{\gamma} A^{-1}(v_\sigma^i)\partial_i f(x),$$

where A^{-1} only acts on v_σ^i by linearity. Secondly, we know that in general the generator of Brownian motion with drift is given by $c^i\partial_i f + \frac{1}{2}D^{ij}\partial_i\partial_j f$. Then, since $L_\varepsilon f_\varepsilon$ has to converge to Brownian motion, the terms associated with ε^0 have to converge to a diffusion

process with possible drift, so we require that

$$\begin{aligned}(c^i\partial_i + \frac{1}{2}D^{ij}\partial_i\partial_j)f(x) &= \frac{1}{2}\left[\kappa\mathrm{Cov}(Z)^{ij} + \lambda v_\sigma^i v_\sigma^j\right]\partial_i\partial_j f(x) + \lambda v_\sigma^i \partial_i g(x,\sigma) + \gamma A h(x,\sigma) \\ &= \frac{1}{2}\left[\kappa\mathrm{Cov}(Z)^{ij} + \lambda v_\sigma^i v_\sigma^j\right]\partial_i\partial_j f(x) - \frac{\lambda^2}{\gamma}v_\sigma^i\partial_i(A^{-1}(v_\sigma^j)\partial_j f(x)) \\ &+ \gamma A h(x,\sigma) \\ &= \frac{1}{2}\left[\kappa\mathrm{Cov}(Z)^{ij} + \lambda v_\sigma^i v_\sigma^j - \frac{2\lambda^2}{\gamma}v_\sigma^i A^{-1}(v_\sigma^j)\right]\partial_i\partial_j f(x) \\ &- \frac{\lambda^2}{\gamma}v_\sigma^i\partial_i A^{-1}(v_\sigma^j)\partial_j f(x) + \gamma A h(x,\sigma),\end{aligned}$$

where it is important to recall that the vector field v_σ has an x dependence and thus the product rule is applied in the last equality. Also note that we still have one degree of freedom, namely h. However, for h to exist, we need to have $h \in D(A)$. And so we have that with respect to the stationary measure ν, it must hold that $\int_\Sigma Ah(x,\varsigma)\nu(\mathrm{d}\varsigma) = 0$. Thus, we find that for h to exist we require the following conditions

$$\begin{aligned}\gamma A h(x,\sigma) = \frac{1}{2}&\left[D^{ij} - \kappa\mathrm{Cov}(Z)^{ij} - \lambda v_\sigma^i v_\sigma^j + \frac{2\lambda^2}{\gamma}v_\sigma^i A^{-1}(v_\sigma^j)\right]\partial_i\partial_j f(x) \\ &+ \left[c^j + \frac{\lambda^2}{\gamma}v_\sigma^i\partial_i A^{-1}(v_\sigma^j)\right]\partial_j f(x) \\ &\Downarrow \\ \int_\Sigma \frac{1}{2}&\left[D^{ij} - \kappa\mathrm{Cov}(Z)^{ij} - \lambda v_\varsigma^i v_\varsigma^j + \frac{2\lambda^2}{\gamma}v_\varsigma^i A^{-1}(v_\varsigma^j)\right]\partial_i\partial_j f(x) \\ &+ \left[c^j + \frac{\lambda^2}{\gamma}v_\varsigma^i\partial_i A^{-1}(v_\varsigma^j)\right]\partial_j f(x)\nu(\mathrm{d}\varsigma) = 0.\end{aligned}$$

Since f is arbitrary, we can select specific components of the equation by taking $f = \pi^i$ or $f = \pi^i\pi^j$. Thus, we find that the above condition can only be fulfilled if the following two conditions hold

$$\forall j \in \{1,...,d\} : \int_\Sigma c^j - v_\varsigma^i\partial_i(-A^{-1})(v_\varsigma^j)\,\nu(\mathrm{d}\varsigma) = 0; \tag{5}$$

$$\begin{aligned}\forall i,j \in \{1,...,d\} : \int_\Sigma &D^{ij} - \kappa\mathrm{Cov}(Z)^{ij} - \lambda v_\varsigma^i v_\varsigma^j - \frac{2\lambda^2}{\gamma}v_\varsigma^i(-A^{-1})(v_\varsigma^j) + \\ &D^{ji} - \kappa\mathrm{Cov}(Z)^{ji} - \lambda v_\varsigma^j v_\varsigma^i - \frac{2\lambda^2}{\gamma}v_\varsigma^j(-A^{-1})(v_\varsigma^i)\,\nu(\mathrm{d}\varsigma) = 0.\end{aligned} \tag{6}$$

Let us first consider equation 5. This equation is associated with the derivative $\partial_j f(x)$, which is the drift term. And so we find that

$$c^j = \int_\Sigma v_\varsigma^i\partial_i(-A^{-1})(v_\varsigma^j)\,\nu(\mathrm{d}\varsigma).$$

Equation 6 fixes the diffusion coefficient. If we recall that the diffusion matrix must be symmetric and that terms $Cov(Z)^{ij}$ and $v_\varsigma^i v_\varsigma^j$ are symmetric, it follows from equation

6 that for all $i, j \in \{1, ..., d\}$,

$$D^{ji} = D^{ij} = \kappa \mathrm{Cov}(Z)^{ij} + \lambda \int_\Sigma v_\varsigma^i v_\varsigma^j \nu(\mathrm{d}\varsigma) + \frac{\lambda^2}{\gamma} \int_\Sigma v_\varsigma^i (-A^{-1})(v_\varsigma^j) + v_\varsigma^j (-A^{-1})(v_\varsigma^i) \nu(\mathrm{d}\varsigma).$$

So overall, we find that $L_\varepsilon f_\varepsilon(x, \sigma) \to c^j \partial_j f(x) + \frac{1}{2} D^{ij} \partial_i \partial_j f(x)$. □

Note that every constant vector field, which is what is used in the previous section, yields $\partial_i(-A^{-1})(v_\varsigma^j) = 0$. So, the drift disappears. Furthermore, for a constant vector field, the diffusion matrix is constant. Thus, in this case, we find that the diffusive limit of the active particle process is Brownian motion with diffusion matrix as in section 4.1.

4.4 Telegrapher Process Scaling

Apart from the diffusive limit, we can also consider another scaling, which does not remove the σ dependence of the process. In this case, we scale time with ε^{-2} and position with ε, but in addition, we also scale the rates λ by ε and γ by ε^2.

Proposition 4.3. *Let $(X_t, \sigma_t)_{t \geq 0}$ be the active particle process. Then, under the above scaling, this process converges to the process with generator*

$$\kappa Cov(Z)^{ij} \partial_i \partial_j f(x, \sigma) + \lambda v_\sigma^i \partial_i f(x, \sigma) + \gamma A f(x, \sigma)$$

Proof. We then get the following generator for the scaled process

$$\begin{aligned} L_\varepsilon f(x, \sigma) = \varepsilon^{-2} \kappa \int_{\mathbb{R}^d} f(x + \varepsilon z, \sigma) - f(x, \sigma) \mu_Z(\mathrm{d}z) \\ + \varepsilon^{-1} \lambda (f(x + \varepsilon v_\sigma, \sigma) - f(x, \sigma)) \\ + \gamma A f(x, \sigma). \end{aligned}$$

If we now Taylor expand f in its first parameter, we find

$$\begin{aligned} L_\varepsilon f(x, \sigma) = \kappa \mathrm{Cov}(Z)^{ij} \partial_i \partial_j f(x, \sigma) \\ + \lambda v_\sigma^i \partial_i f(x, \sigma) \\ + \gamma A f(x, \sigma), \end{aligned}$$

which converges as $\varepsilon \to 0$. □

Looking at the generator, we see that the process consists of Brownian motion with σ-dependent drift. This can be viewed as Brownian motion on a collection of copies of $\mathbb{R}^d$ indexed by Σ, where the drift of the Brownian motion depends on which copy of $\mathbb{R}^d$ the particle resides in. This process is actually quite relevant, since – as we will see later on – it is closely tied to the Dirac equation.

4.5 Examples

We will now consider two more specific examples of active particle processes to make some of the above notions more concrete.

Example 7. We begin by considering a run and tumble motion in $\mathbb{R}^3$. We, then, have the unit ball $\mathbb{S}^2$ as internal state. Then with a rate γ, we tumble by selecting an entirely new random vector from $\mathbb{S}^2$ with uniform probability. Note that we then take the process acceleration γ, which introduced in our model to be 1. Our velocity function will be $v(\sigma) = \sigma$. Thus, our particle will walk in the direction of its internal state. Since the random walk part is completely independent of the active movement, we can leave it out without loss of generality. We will now compute the diffusion matrix of this particle in the diffusive limit. Firstly, let us consider

$$\lambda \int_{\mathbb{S}^2} v_\varsigma^i v_\varsigma^j \nu(\mathrm{d}\varsigma) = \lambda \int_{\mathbb{S}^2} \varsigma^i \varsigma^j \nu(\mathrm{d}\varsigma).$$

Now, observe that unless $i = j$, there is no correlation between the two components of ς. In fact, if we fix ς^i, we immediately note that there is a perfectly symmetric circle of values for ς^j. So if $i \neq j$, we find that the above integral is zero.

$$\begin{aligned}
\lambda \int_{\mathbb{S}^2} v_\varsigma^i v_\varsigma^j \nu(\mathrm{d}\varsigma) &= \lambda \int_{\mathbb{S}^2} (\varsigma^i)^2 \nu(\mathrm{d}\varsigma) \delta^{ij} \\
&= \lambda \int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\sin\theta)^2 \cos\theta \, \mathrm{d}\theta \, \mathrm{d}\phi \, \delta^{ij} \\
&= 2\pi\lambda \int_{-1}^{1} x^2 \, \mathrm{d}x \, \delta^{ij} \\
&= \frac{4}{3}\pi\lambda\delta^{ij},
\end{aligned}$$

where the integral was switched to spherical coordinates and the substitution $x = sin\theta$ was used. Now, for the next part of the diffusion coefficient, we need to evaluate $-A^{-1}$. To do so, we will use the simple observation that after one jump $\mathbb{E}[\sigma] = 0$, as we choose a new vector uniformly from $\mathbb{S}^2$. Now, observe that the probability of jumping before a time t is $1 - e^{-\gamma t}$, so the probability of not jumping is $e^{-\gamma t}$. Therefore, we find that

$$\begin{aligned}
-A^{-1}\varsigma^i &= \int_0^\infty S_t \varsigma^i \, \mathrm{d}t \\
&= \int_0^\infty \varsigma^i \cdot e^{-\gamma t} + 0 \cdot (1 - e^{\gamma t}) \, \mathrm{d}t \\
&= \varsigma^i \int_0^\infty e^{-\gamma t} \, \mathrm{d}t \\
&= \frac{\varsigma^i}{\gamma}.
\end{aligned}$$

Then,using our previous calculation, we see that

$$\begin{aligned}
\lambda^2 \int_\Sigma v_\varsigma^i (-A^{-1})(v_\varsigma^j) + v_\varsigma^j (-A^{-1})(v_\varsigma^i) \nu(\mathrm{d}\varsigma) &= \frac{2\lambda^2}{\gamma} \int_\Sigma \varsigma^i \varsigma^j \nu(\mathrm{d}\varsigma) \\
&= \frac{8\pi}{3} \frac{\lambda^2}{\gamma} \delta^{ij}
\end{aligned}$$

So our diffusion matrix is then given by $D^{ij} = (\frac{4}{3}\pi\lambda + \frac{8\pi}{3}\frac{\lambda^2}{\gamma})\delta^{ij}$. △

Example 8. The following example comes from [1]. In this case, we consider an $\mathbb{R}$ valued active particle process with as an internal state an Ornstein-Uhlenbeck process. Namely, the process satisfying

$$\mathrm{d}\sigma_t = -a\sigma_t \,\mathrm{d}t + b\,\mathrm{d}B_t\,,$$

with B_t is independent standard Brownian motion. We can view this as the process σ_t satisfying the differential equation $y' = -ay$ with a noise component introduced by B_t. Furthermore, we take the velocity function $v : \mathbb{R} \to \mathbb{R}$ defined by $\sigma \mapsto \sigma$.

We now want to find the diffusive limit of the above process. The internal state process has stationary distribution $\nu \sim N(0, \frac{b^2}{2a})$. Moreover, we then have the generator

$$A = -a\sigma\frac{\mathrm{d}}{\mathrm{d}\sigma} + \frac{b^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}\sigma^2}.$$

Now, the function $w(\sigma) = \frac{\sigma}{a}$ is contained in $D(A)$, since $\int w(\varsigma)\nu(\mathrm{d}\varsigma) = 0$. Moreover, $-Aw(\sigma) = \sigma = v(\sigma)$. Thus, $w = -A^{-1}v$. Thus we find that

$$\lambda\int_{\mathbb{R}} \varsigma^2\,\nu(\mathrm{d}\varsigma) = \frac{\lambda b^2}{2a}$$
$$\frac{2\lambda^2}{\gamma}\int_{\mathbb{R}} \frac{\varsigma^2}{a}\,\nu(\mathrm{d}\varsigma) = \frac{\lambda^2}{\gamma}\frac{b^2}{a^2}.$$

So overall, we find that diffusion coefficient of the diffusive limit of this process would be

$$D = \frac{\lambda b^2}{2a} + \frac{\lambda^2}{\gamma}\frac{b^2}{a^2}.$$

△

5 Active Particles on Riemannian Manifolds

We would now like to extend the active particle process to a Riemannian manifold. However, this first raises the question what is Brownian motion on a Riemannian manifold, since we can no longer exploit the vector space structure of $\mathbb{R}^n$ to define Brownian motion. This question was solved in a paper by Jørgensen [2], where he defines Brownian motion as the diffusive limit of a geodesic random walk. Jørgensen's paper and the introductory chapters of [3] form the basis of the first two sections of the chapter.

We will start of by exploring what exactly a geodesic random walk is and then continue on to the diffusive limit of this process. This limit is then used to define Brownian motion. Furthermore, some more considerations about the nature of this Brownian motion will be given. With these ideas explored, we will move onto defining the active particle motion on Riemannian manifolds and thereafter, show that analogously to chapter 4, this process can converge to Brownian motion in the diffusive limit. In addition, we will show that a limit that conforms with the telegrapher process also exists.

Throughout this chapter, we will work in a compact manifold M with metric g. For a point $p \in M$, we denote its tangent space by T_pM. Furthermore, we use $\mathrm{d}f$ to denote the exterior derivative of f and $\mathrm{d}^2 f$ to denote the Laplace-Beltrami operator. In local coordinates around some point $p \in M$ with orthonormal basis ∂_i of T_pM, the exterior derivative $\mathrm{d}f$ is given by $\mathrm{d}f\,(X) = \xi^i \partial_i f$ with $X = \xi^i \partial_i$. And in these same local coordinates, the Laplace-Beltrami operator is given by $\mathrm{d}^2 f\,(X \otimes Y) = (\partial_i \partial_j f - \Gamma_{ij}^k \partial_k f)\xi^i \eta^j$, where $X = \xi^i \partial_i, Y = \eta^j \partial_j \in T_pM$ and Γ_{ij}^k denotes the Christoffel symbols of the manifold. Moreover, for all of the following sections, it is assumed that there is some joint underlying probability space $(\Omega, \mathscr{A}, \mu)$.

5.1 Geodesic Random Walk

In $\mathbb{R}^n$, we can view a random walk as a sum of random independent increments. So, we could say that if we start out at some point $x \in \mathbb{R}^d$ and have a sequence of random vectors Z_i, $X_0 = x$ and $X_n = X_0 + \sum_{i=1}^n Z_i$ together define our random walk.

In the case of manifolds, we cannot simply add a tangent vector to a point. This operation is not well defined. However, we do know that for every $p \in M$ and $v \in T_pM$, there exists a geodesic $\gamma : \mathbb{R} \supseteq (a, b) \to M$ such that $0 \in [a, b]$, $\gamma(0) = p$ and $v = \dot{\gamma}(0) = \frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} \gamma(t)$. Recall that a geodesic is the equivalent of a straight line inside our manifold. Now, we will assume that $[0, 1] \subseteq [a, b]$. This is not the case in general, but also not a too restrictive constriction as many import examples of submanifold in $\mathbb{R}^d$ have this property, primarily the sphere and torus.

The idea is to now define an increment associated with v by following a geodesic for one unit of time. This leads us to the Riemannian exponential map $\exp : TM \to M$ defined as $(p, v) \mapsto \exp_p(v) = \gamma(1)$. One especially useful result for the Riemannian exponential map is a special version of Taylor's theorem on manifolds.

Theorem 5.1. *Let $f \in C^2(M)$, $p \in M$ and $v \in T_pM$. Then,*

$$f(\exp_p(v)) = f(p) + \mathrm{d}f_p\,(v) + \frac{1}{2}\,\mathrm{d}^2 f_p\,(v \otimes v) + o(\|v\|^2),$$

where $\|v\|$ denotes $\sqrt{g(v, v)}$.

Now, we can define a geodesic random walk as a concatenation of Riemannian exponential maps, which essentially comes down to following a path of random geodesics.

Definition 5.2. *Let $p_0 \in M$. Then, we call the collection $(X_n)_{n\geq 0}$ of M-valued random variables a geodesic random walk with increments $(Z_n)_{n\geq 0}$ starting at p_0 if*

(i) $X_0 = p_0$;

(ii) *for all $n \geq 0$, $Z_n \in T_{X_n}M$;*

(iii) *for all $n \geq 0$, $X_{n+1} = \exp_{X_n}(Z_n)$.*

Remark 2. Since in $\mathbb{R}^d$ the Riemannian exponential map reduces to $\exp_p(v) = p + v$, we have that the above definition in $\mathbb{R}^d$ reduces exactly to the random walk as defined in section 3.1. △

Now, in order for a geodesic random walk to be a Markov process, we need that the distribution of an increment Z_n only depends on the location X_n. We do this by associating to each point $p \in M$ a random variable Z_p with measure μ_p on T_pM. So, then, the one time step transition kernel P acts on $f \in C(M)$ as

$$Pf(p) = \int_{T_pM} f(\exp_p(z))\, \mu_p(\mathrm{d}z).$$

Now, to transform the above process into a continuous time process $(X_t)_{t\geq 0}$, we transform it into a jump process, which jumps at a rate κ. So, then using example 2, we find that the process has generator

$$Lf(p) = \kappa(P - I)f(p) = \kappa \int_{T_pM} [f(\exp_p(z)) - f(p)]\, \nu_p(\mathrm{d}z).$$

5.2 Scaling and Brownian Motion

In $\mathbb{R}^d$, when we have collection of random increments $(Z_n)_{n\geq 0}$ and we scale space by ε, we consider $\varepsilon \sum_0^m Z_n$. However, in manifolds, we have replaced these sums with exponential maps, so we cannot scale the entire trajectory. Now, note that we can also write the scaled sum of the trajectory as $\sum_0^m (\varepsilon Z_n)$. Similarly, in order to scale space in a manifold, we scale all the individual tangent vectors of our path, which is possible as T_pM is a vector space over $\mathbb{R}$.

Furthermore, we want to make a special consideration for the case that our random variables Z_n have a drift, which was ignored in section 3.3. We know that $\frac{1}{\sqrt{n}} \sum_{k=0}^n Z_k$ diverges as $n \to \infty$ if Z_k exhibits a drift. To compensate for this, we switch to scaling $m + \frac{1}{\sqrt{n}} \sum_{k=0}^n (Z_k - m)$, where $m = \mathbb{E}[Z_0]$. In a similar, way we can compensate for drift in a scaled random walk.

Definition 5.3. *Let $\varepsilon > 0$ and $p_0 \in M$. And let $(X_n)_{n\geq 0}$ be a geodesic random walk with increments $(Z_n)_{n\geq 0}$ starting at p_0. Then, we call $(X_n^\varepsilon)_{n\geq 0}$ a ε-rescaled random walk with increments $(Z_n)_{n\geq 0}$ starting at p_0 if*

(i) $X_0^\varepsilon = p_0$;

(ii) *for all* $n \geq 0$, $Z_n \in T_{X_n^\varepsilon} M$;

(iii) *for all* $n \geq 0$, $X_{n+1}^\varepsilon = \exp_{X_n^\varepsilon}(\varepsilon^2 \mathbb{E}[Z_n] + \varepsilon(Z_n - \mathbb{E}[Z_n]))$.

Then, we can define diffusive scaling exactly as in section 3.3.

Definition 5.4. *Let* $(X_t)_{t\geq 0}$ *be a geodesic random walk. Then, the diffusive scaling of this process is defined as* $\lim_{\varepsilon\to 0} X^\varepsilon_{\varepsilon^{-2}t}$ *if it exists. Here, this limit is in the sense of weak convergence in the path space.*

Consider the continuous random walk $(X_t)_{t\geq 0}$ from the previous section. Let m_p denote $\mathbb{E}[Z_p]$, $m_p^2 = \mathbb{E}[Z_p \otimes Z_p]$ and $\sigma_p^2 = m_p^2 - m_p \otimes m_p$. Then, under diffusive scaling, we get the process $(X^\varepsilon_{\varepsilon^{-2}t})_{t\geq 0}$ with the generator

$$L_\varepsilon f(p) = \varepsilon^{-2}\kappa(P_\varepsilon - I)f(p) = \varepsilon^{-2}\kappa \int_{T_pM} [f(\exp_p(\varepsilon^2 m_p + \varepsilon(z - m_p))) - f(p)]\, \mu_p(\mathrm{d}z).$$

Now, applying Taylor's theorem for exponential maps, we find that

$$\begin{aligned}
L_\varepsilon f(p) &= \varepsilon^{-2}\kappa \int_{T_pM} [\mathrm{d}f\,(\varepsilon^2 m_p + \varepsilon[z - m_p]) \\
&\quad + \frac{1}{2}\,\mathrm{d}^2 f\,((\varepsilon^2 m_p + \varepsilon[z - m_p]) \otimes (\varepsilon^2 m_p + \varepsilon[z - m_p])) + o(\varepsilon^2)]\, \mu_p(\mathrm{d}z) \\
&\overset{*}{=} \kappa\,\mathrm{d}f\,(m_p) + \kappa\varepsilon^{-1}\,\mathrm{d}f\left(\int_{T_pM} z\,\mu_p(\mathrm{d}z) - m_p\right) \\
&\quad + \frac{\kappa}{2}\,\mathrm{d}^2 f\left(\int_{T_pM} (z - m_p) \otimes (z - m_p)\,\mu_p(\mathrm{d}z)\right) + o(1) \\
&\overset{**}{=} \kappa\,\mathrm{d}f\,(m_p) + 0 + \frac{\kappa}{2}\,\mathrm{d}^2 f\left(\int_{T_pM} z \otimes z\,\mu_p(\mathrm{d}z) - m_p \otimes m_p\right) + o(1) \\
&= \kappa\,\mathrm{d}f\,(m_p) + \frac{\kappa}{2}\,\mathrm{d}^2 f\,(\sigma_p^2) + o(1),
\end{aligned} \tag{7}$$

where at $*$, the linearity of $\mathrm{d}f$ and bilinearity of $\mathrm{d}^2 f$ was used and at $**$, it was used that $\int_{T_pM} z \otimes m_p\,\mu_p(\mathrm{d}z) = m_p \otimes m_p$. Then, we find that as $\varepsilon \to 0$, $\varepsilon X_{\varepsilon^{-2}t}$ converges to the process with the generator

$$\kappa\,\mathrm{d}f\,(m_p) + \frac{\kappa}{2}\,\mathrm{d}^2 f\,(\sigma_p^2).$$

This will actually be our definition for Brownian motion on manifolds. However, before we actually define this, we have one more consideration. In $\mathbb{R}^d$, we have that Brownian motion has a constant diffusion matrix and drift. This is in fact associated with the fact all the increments are identically distributed in $\mathbb{R}^d$.

To extend the notion of a constant tensor field to manifolds, we need something called parallel transport. The idea is that we will move a vector along a curve without changing its direction relative to the curve. More specifically, we will transport our distributions along geodesics.

Assume we have two point $p, q \in M$ and an arbitrary path of geodesics between them. Now, let τ_{pq} denote the parallel transport along this path p to q. Then, we

require that $\mu_q = \tau_{pq}\mu_p$. We say that our measures are invariant under parallel transport. Furthermore, we have that a measure μ_p in one point $p \in M$ defines the measure μ_q for all $q \in M$. In addition, it holds that $m_q = \tau_{pq} m_p$ and $\sigma_q^2 = \tau_{pq}\sigma_p^2$.

Definition 5.5. *We call a manifold valued process $(X_t)_{t\geq 0}$ Brownian motion, if it has the generator*

$$Lf(p) = \mathrm{d}f\,(a_p) + \frac{1}{2}\,\mathrm{d}^2 f\,(b_p),$$

where $a_p \in T_pM$ and $b_p \in T_pM \otimes T_pM$. Furthermore, we require that b_p is a symmetric, positive semi-definite tensor field and that both tensor fields are invariant under parallel transport along geodesics.

Remark 3. In $\mathbb{R}^d$, we have that being invariant under parallel transport corresponds to being constant. And thus, we see that the above definition reduces to the definition $\mathbb{R}^d$-value Brownian motion with drift. △

Now, as a relaxation of the condition that m_p and σ_p^2 are invariant under parallel transport along geodesics, we could take the condition that m_p and σ_p^2 are smoothly varying tensor fields. Then, we see that we obtain a manifold equivalent for $\mathbb{R}^d$-valued Brownian motion with varying diffusion matrix as encountered in section 4.3.

5.3 Active Particles

We will now consider a manifold-valued active particle process. In fact not a lot has changed from section 4.1, we add an internal state space $(\Sigma, \mathcal{S})$ to the compact manifold M. So that we consider a process $(X_t, \sigma_t)_{t\geq 0}$ in $M \times \Sigma$. Furthermore, we have a velocity map $v : \Sigma \times M \to TM$, which specifies for each $\sigma \in \Sigma$, a smooth vector field v_σ.

The internal state process $(\sigma_t)_{t\geq 0}$ is exactly identical to the internal state process defined in section 4.1. So, it has a generator A and stationary measure ν. Furthermore, we again assume A^{-1} exists.

Now, the location process $(X_t)_{t\geq 0}$ changes in a natural way from $\mathbb{R}^d$ to M. We still have that $(X_t)_{t\geq 0}$ evolves in accordance with two independent jump processes. Firstly, a random walk as in section 5.1, which has the generator

$$\kappa \int_{T_pM} [f(\exp_p(z)) - f(p)]\,\nu_p(\mathrm{d}z).$$

Moreover, we assume that the measure ν_p is invariant under parallel transport, so that this part of the process converges to Brownian motion with a constant diffusion matrix.

Secondly, we jump along the velocity vector field v_σ at rate λ depending on the current internal state $\sigma \in \Sigma$. From this, it immediately follows that the generator is given by

$$\lambda[f(\exp_p(v_\sigma(p)), \sigma) - f(p, \sigma)].$$

In addition, for every $p \in M$, we will assume that $\mathbb{E}_\nu[v(\sigma)] = \int_\Sigma v(\varsigma, p)\,\nu(\mathrm{d}\varsigma) = 0$, $v(\cdot, p) \in L^2(\nu)$ and $A^{-1}v(\cdot, p)$ exists.

So once again, an active particle process is a random walk with an additional drift component, which is dictated by the internal state of the particle. Now, the generator L

for the complete active particle process is quite simply the sum of the generators of the individual processes as per example 2. So, we find

$$\begin{aligned} Lf(p,\sigma) = \kappa \int_{T_pM} [f(\exp_p(z)) - f(p)]\, \nu_p(\mathrm{d}z) \\ + \lambda[f(\exp_p(v_\sigma(p)), \sigma) - f(p,\sigma)] \\ + \gamma A f(p,\sigma) \end{aligned}$$

We will also need the generator L_ε of the process $(\varepsilon X_{\varepsilon^{-2}t}, \sigma_{\varepsilon^{-2}t})_{t\geq 0}$, which is the diffusive scaling of the above process and is given by

$$\begin{aligned} L_\varepsilon f(p,\sigma) &= \varepsilon^{-2}\kappa \int_{T_pM} [f(\exp_p(\varepsilon^2 m_p + \varepsilon(z-m_p)), \sigma) - f(p,\sigma)]\, \nu_p(\mathrm{d}z) \\ &\quad + \varepsilon^{-2}\lambda[f(\exp_p\{\mathbb{E}_\nu[v_\sigma(p)] + \varepsilon(v_\sigma(p) - \mathbb{E}_\nu[v_\sigma(p)])\}, \sigma) - f(p,\sigma)] \\ &\quad + \varepsilon^{-2}\gamma A f(p,\sigma) \\ &= \varepsilon^{-2}\kappa \int_{T_pM} [f(\exp_p(\varepsilon^2 m_p + \varepsilon(z-m_p)), \sigma) - f(p,\sigma)]\, \nu_p(\mathrm{d}z) \\ &\quad + \varepsilon^{-2}\lambda[f(\varepsilon v_\sigma(p), \sigma) - f(p,\sigma)] \\ &\quad + \varepsilon^{-2}\gamma A f(p,\sigma) \end{aligned} \tag{8}$$

We will now proof that the above process converges to Brownian motion in the same way as was done in section 4.3.

Theorem 5.6. *Let $(X_t, \sigma_t)_{t\geq 0}$ be an manifold-valued active particle process with a collection of smooth vector fields as velocity function. Then, the diffusive limit of this process converges to the process with generator $\mathrm{d}f\,(a_p) + \frac{1}{2}\,\mathrm{d}^2 f\,(b_p)$ on $C^2(M)$, where*

$$a_p = \kappa m_p + \int_\Sigma \nabla_{v_\varsigma}(-A^{-1}v_\varsigma)\, \nu(\mathrm{d}\varsigma)$$

with ∇ denoting the Levi-Civita connection and

$$b_p = \kappa\sigma_p^2 + \lambda \int_\Sigma v_\varsigma^2\, \nu(\mathrm{d}\varsigma) + \frac{\lambda^2}{\gamma} \int_\Sigma [v_\varsigma \otimes (-A^{-1}v_\varsigma) + (-A^{-1}v_\varsigma) \otimes v_\varsigma]\, \nu(\mathrm{d}\varsigma)$$

are smoothly varying coefficients.

In the following, proof we will assume there exists a chart around $p \in M$ such that TM has an local smooth orthonormal vector field $(\partial_i)_{1\leq i\leq d}$ around p with dual basis $(\mathrm{d}x^i)_{1\leq i\leq d}$, where d is the dimension of the manifold M.

Proof. Assume we have an arbitrary function f in the domain of Brownian motion $C^2(M)$. Then, we consider the sequence of function $f_\varepsilon(p,\sigma) = f(p) + \varepsilon g(x,\sigma) + \varepsilon^2 h(p,\sigma)$, where g, h are to be determined. Now, consider the generator L_ε from equation 8 and apply it to f_ε. Then, using the manifold Taylor expansion for exponential functions and leaving

out the location dependence of v_σ for readability, we see that

$$\begin{aligned}
L_\varepsilon f_\varepsilon(p,\sigma) &= \varepsilon^{-2}\kappa \int_{T_pM} [f_\varepsilon(\exp_p(\varepsilon^2 m_p + \varepsilon(z-m_p))) - f_\varepsilon(p,\sigma)]\,\nu_p(\mathrm{d}z) \\
&\quad + \varepsilon^{-2}\lambda[f_\varepsilon(\exp_p(\varepsilon v_\sigma),\sigma) - f_\varepsilon(p,\sigma)] \\
&\quad + \varepsilon^{-2}\gamma A f(p,\sigma) \\
&= \kappa\,\mathrm{d}f_\varepsilon\,(m_p,\sigma) + \frac{\kappa}{2}\,\mathrm{d}f_\varepsilon\,(\sigma_p^2,\sigma) \\
&\quad + \varepsilon^{-1}\lambda\,\mathrm{d}f_\varepsilon\,(v_\sigma,\sigma) + \frac{\lambda}{2}\,\mathrm{d}^2 f_\varepsilon\,(v_\sigma \otimes v_\sigma,\sigma) \\
&\quad + \varepsilon^{-2}\gamma A f(p,\sigma) + o(1),
\end{aligned}$$

where in the last equality, we used that the random walk generator becomes equation 7 and that $\mathrm{d}f_\varepsilon$ and $\mathrm{d}^2 f_\varepsilon$ are linear and bilinear respectively. Furthermore, we have collected all vanishing terms in $o(1)$. From now on, we will denote $v_\sigma \otimes v_\sigma$ by v_σ^2.

If we now substitute in $f_\varepsilon(p,\sigma) = f(p) + \varepsilon g(x,\sigma) + \varepsilon^2 h(p,\sigma)$, immediately recollect all vanishing terms in $o(1)$ and note that $Af(p) = 0$, we find that

$$\begin{aligned}
L_\varepsilon f_\varepsilon(p,\sigma) &= \kappa\,\mathrm{d}f\,(m_p) + \frac{1}{2}\,\mathrm{d}f\,(\kappa\sigma_p^2 + \lambda v_\sigma^2) + \varepsilon^{-1}\lambda\,\mathrm{d}f\,(v_\sigma) \\
&\quad + \lambda\,\mathrm{d}g\,(v_\sigma,\sigma) + \varepsilon^{-1}\gamma A g(p,\sigma) + \gamma A h(p,\sigma) + o(1),
\end{aligned}$$

If we want the above expression to converge to a generator of Brownian motion for all f as $\varepsilon \to 0$, we need that terms associated with ε^{-1} vanish. Thus,

$$\lambda\,\mathrm{d}f\,(v_\sigma) + \gamma A g(p,\sigma) = 0 \implies g(p,\sigma) = \frac{\lambda}{\gamma}(-A^{-1})(\mathrm{d}f\,(v_\sigma)).$$

In fact, we can pull A^{-1} inside $\mathrm{d}f$, which follows directly from changing to local coordinates:

$$\begin{aligned}
g(p,\sigma) &= \frac{\lambda}{\gamma}(-A^{-1})(\mathrm{d}f\,(v_\sigma)) \\
&= \frac{\lambda}{\gamma}(-A^{-1})(\partial_i f v_\sigma^j\,\mathrm{d}x^i\,(\partial_j)) \\
&= \frac{\lambda}{\gamma}(-A^{-1})(\partial_i f v_\sigma^i) \\
&= \frac{\lambda}{\gamma}\partial_i f(-A^{-1})v_\sigma^i \\
&= \frac{\lambda}{\gamma}\partial_i f\,\mathrm{d}x^i\,(-A^{-1}(v_\sigma^j\partial_j)) \\
&= \frac{\lambda}{\gamma}\,\mathrm{d}f\,(-A^{-1}v_\sigma).
\end{aligned}$$

Now, we know that a generator of Brownian motion with drift is of the form $\mathrm{d}f\,(a_p) + \frac{1}{2}\,\mathrm{d}^2 f\,(b_p)$, where $a_p \in T_pM$ and $b_p \in T_pM \otimes T_pM$. Then, since we want $L_\varepsilon f_\varepsilon$ to converge

to Brownian motion, the terms associated with ε^0 have to converge to something of this form, so we require that

$$\begin{aligned}\mathrm{d}f\,(a_p) + \frac{1}{2}\,\mathrm{d}^2 f\,(b_p) &= \kappa\,\mathrm{d}f\,(m_p) + \frac{1}{2}\,\mathrm{d}f\,(\kappa\sigma_p^2 + \lambda v_\sigma^2) + \lambda\,\mathrm{d}g\,(v_\sigma, \sigma) + \gamma A h(p, \sigma) \\ &= \kappa\,\mathrm{d}f\,(m_p) + \frac{1}{2}\,\mathrm{d}f\,(\kappa\sigma_p^2 + \lambda v_\sigma^2) + \frac{\lambda^2}{\gamma}\,\mathrm{d}(\mathrm{d}f\,(-A^{-1}v_\sigma))\,(v_\sigma) \\ &\quad + \gamma A h(p, \sigma).\end{aligned}$$

Now, the term $\frac{\lambda^2}{\gamma}\,\mathrm{d}(\mathrm{d}f\,(-A^{-1}v_\sigma))\,(v_\sigma)$ warrants further inspection. Firstly, note that this term is not the exterior derivative of the 1-form $\mathrm{d}f$, but rather the exterior derivative of the 0-form $\mathrm{d}f\,(-A^{-1}v_\sigma)$. By going to local coordinates, we see using the Leibniz rule that

$$\begin{aligned}\mathrm{d}(\mathrm{d}f\,(-A^{-1}v_\sigma))\,(v_\sigma) &= \mathrm{d}(\partial_i f(-A^{-1}v_\sigma^i))\,(v_\sigma) \\ &= (-A^{-1}v_\sigma^i)\,\mathrm{d}(\partial_i f)\,(v_\sigma) + \partial_i f\,\mathrm{d}(-A^{-1}v_\sigma^i)\,(v_\sigma) \\ &= v_\sigma^j(-A^{-1}v_\sigma^i)\partial_j\partial_i f + v_\sigma^j\partial_j(-A^{-1}v_\sigma^i)\partial_i f \\ &= (\partial_j\partial_i f - \Gamma_{ij}^k\partial_k f)v_\sigma^j(-A^{-1}v_\sigma^i) \\ &\quad + v_\sigma^j(-A^{-1}v_\sigma^i)\Gamma_{ij}^k\partial_k f + v_\sigma^j\partial_j(-A^{-1}v_\sigma^i)\partial_i f.\end{aligned}$$

This first term is an expression for the Laplace-Beltrami operator in local coordinates. Regarding the second term, recall that for two vector field $u = u^i\partial_i, v = v^j\partial_j$, we have

$$\begin{aligned}\nabla_v u &= v^j\nabla_{\partial_j}(u^i\partial_i) \\ &= v^j u^i\nabla_{\partial_j}\partial_i + v^j\nabla_{\partial_j}(u^i)\partial_i \\ &= v^j u^i\Gamma_{ij}^k\partial_k + v^j\partial_j(u^i)\partial_i,\end{aligned}$$

where ∇ denotes the Levi-Civita connection. Using this result, we then find that

$$\begin{aligned}\mathrm{d}(\mathrm{d}f\,(-A^{-1}v_\sigma))\,(v_\sigma) &= \mathrm{d}^2 f\,(v_\sigma \otimes (-A^{-1}v_\sigma)) + \nabla_{v_\sigma}(-A^{-1}v_\sigma)f \\ &= \mathrm{d}^2 f\,(v_\sigma \otimes (-A^{-1}v_\sigma)) + \mathrm{d}f\,(\nabla_{v_\sigma}(-A^{-1}v_\sigma)),\end{aligned}$$

So overall, we have

$$\begin{aligned}\mathrm{d}f\,(a_p) + \frac{1}{2}\,\mathrm{d}^2 f\,(b_p) = \kappa\,\mathrm{d}f\,(m_p) + \frac{1}{2}\,\mathrm{d}f\,(\kappa\sigma_p^2 + \lambda v_\sigma^2 + \frac{2\lambda^2}{\gamma}v_\sigma \otimes (-A^{-1}v_\sigma)) \\ + \mathrm{d}f\,(\nabla_{v_\sigma}(-A^{-1}v_\sigma)) + \gamma A h(p, \sigma).\end{aligned}$$

Now, we once again have one degree of freedom left, namely h, with the requirement that $h \in D(A)$. And thus with respect to the stationary measure ν, it must hold that $\int_\Sigma Ah(p, \varsigma)\,\nu(\varsigma) = 0$. Thus, for h to exists, we require that

$$\begin{aligned}\int_\Sigma \mathrm{d}f\,(a_p - \kappa m_p - \nabla_{v_\varsigma}(-A^{-1}v_\varsigma))\,\nu(\mathrm{d}\varsigma) \\ + \int_\Sigma \frac{1}{2}\,\mathrm{d}^2 f\,(b_p - \kappa\sigma_p^2 - \lambda v_\varsigma^2 - \frac{2\lambda^2}{\gamma}v_\varsigma \otimes (-A^{-1}v_\varsigma))\,\nu(\mathrm{d}\varsigma) = 0.\end{aligned}$$

As f is arbitrary, we can choose it such that $\mathrm{d}^2 f$ vanishes, namely taking $f = x^i$ locally, so we find

$$a_p = \kappa m_p + \int_\Sigma \nabla_{v_\varsigma}(-A^{-1}v_\varsigma)\,\nu(\mathrm{d}\varsigma)$$

Now, for the part associated with $\mathrm{d}^2 f$, observe that this operator is symmetric. So, even though we can select parts of the equation by choosing $f = x^i x^j$ locally, we cannot distinguish the left and the right side of a tensor product. Note that σ_p^2 and v_ς^2 are symmetric anyways, so we find that

$$b_p = \kappa\sigma_p^2 + \lambda \int_\Sigma v_\varsigma^2\,\nu(\mathrm{d}\varsigma) + \frac{\lambda^2}{\gamma}\int_\Sigma [v_\varsigma \otimes (-A^{-1}v_\varsigma) + (-A^{-1}v_\varsigma)\otimes v_\varsigma]\,\nu(\mathrm{d}\varsigma).$$

And so overall, we find that $L_\varepsilon f_\varepsilon(p,\sigma) \to \mathrm{d}f\,(a_p) + \frac{1}{2}\,\mathrm{d}^2 f\,(b_p)$. □

Now, it is apriori not clear if taking the vector field v_σ invariant under parallel transport along geodesics for every $\sigma \in \Sigma$ will ensure that a_p and b_p are invariant themselves. Nonetheless, as v_σ is a collection of smooth vector fields, we have that a_p and b_p are smooth tensor fields. So, we can certainly view the limit of the active particle process as Brownian motion with smoothly varying drift and diffusion.

Furthermore, we see that the actual result from the above theorem is completely analogous to the results from section 4.3. If we consider the above generator on $\mathbb{R}^d$, we get back exactly our results from this section. So the diffusive limit of the manifold-valued active particle process is consistent with the diffusive limit of the $\mathbb{R}^d$-valued active particle process.

5.4 Telegrapher Process Scaling

Apart from the diffusive limit, we can once again consider the telegrapher scaling as in section 4.4. Recall that this scaling does not remove the σ dependence of the process,

Proposition 5.7. *Let $(X_t, \sigma_t)_{t\geq 0}$ be the manifold-valued active particle process. Then, under the above scaling, this process converges to the process with generator*

$$\mathrm{d}f\,(\kappa m_p + \lambda v_\sigma, \sigma) + \kappa\,\mathrm{d}^2 f\,(\sigma_p^2, \sigma) + \gamma A f(x,\sigma).$$

Proof. We then get the following generator for the scaled process

$$\begin{aligned} L_\varepsilon f(p,\sigma) = {}& \varepsilon^{-2}\kappa \int_{T_pM} f(\exp_p(\varepsilon^2 m_p + \varepsilon(z-m_p)),\sigma) - f(p,\sigma)\,\mu_Z(\mathrm{d}z) \\ & + \varepsilon^{-1}\lambda(f(\exp_p(\varepsilon v_\sigma),\sigma) - f(p,\sigma)) \\ & + \gamma A f(p,\sigma). \end{aligned}$$

If we now apply the Taylor expansion of the exponential map in manifolds to f, we find

$$\begin{aligned} L_\varepsilon f(x,\sigma) = {}& \kappa\,\mathrm{d}f\,(m_p,\sigma) + \kappa\,\mathrm{d}^2 f\,(\sigma_p^2,\sigma) \\ & + \lambda\,\mathrm{d}f\,(v_\sigma,\sigma) \\ & + \gamma A f(x,\sigma) + o(1), \end{aligned}$$

which converges as $\varepsilon \to 0$. □

Looking at the generator, we see that the process consists of Brownian motion with drift, where the exact drift depends on the current internal state.

6 Quantum Diffusion Monte Carlo

Up to now, we have mainly considered stochastic processes in a pure mathematical sense. However, they are widely used throughout science in a plethora of applications. One such field of applications is in quantum mechanics. Here there exists a link between Brownian motion and the Schrödinger equation. Moreover, this link can be exploited to create a numerical method for finding ground-states of quantum systems. This numerical method is called diffusion Monte Carlo. The main advantage of this method over traditional numerical integration schemes is that the error is independent of the number of dimensions, namely the error is $o(\frac{1}{\sqrt{N}})$, where N is the number of sample points. Whilst for a d dimensional problem, the error of a typical grid based solution is $o(\frac{1}{N^{1/d}})$ [14], where N is the number of sample points. Thus, we notice that if $d > 2$, Monte Carlo starts getting advantageous. The current section is largely based on chapter 12 of [15].

6.1 Solving the Schrödinger equation in imaginary time

The Schrödinger equation is one of the most famous equations in physics. Generally, when it is considered with some time independent potential, it has the following form:

$$i\hbar\frac{\partial}{\partial t}\psi(\mathbf{r},t) = -\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r},t) + V(\mathbf{r})\psi(\mathbf{r},t).$$

We would like to solve this equation in general, at least formally. In order to do so, we will transform the above equation into the form $\frac{\mathrm{d}}{\mathrm{d}\tau}\rho = L\rho$, where L is the generator of a semigroup.

Firstly, set $\tau = it$, then $\frac{\partial}{\partial t} = \frac{\partial \tau}{\partial t}\frac{\partial}{\partial \tau} = i\frac{\partial}{\partial \tau}$ and say that $\rho(\mathbf{r},\tau) = \psi(\mathbf{r},-i\tau)$. Thus, if we then take our units such that $m = \hbar = 1$; take $K = -\frac{1}{2}\nabla^2$; and hide the position dependence of V, we find

$$\frac{\partial}{\partial \tau}\rho(\mathbf{r},\tau) = -(K+V)\rho(\mathbf{r},\tau).$$

To solve our original equation, it now suffices to solve for ρ. Now, note that $-(K+V)$ is a linear operator. In fact, K is a generator of Brownian motion and we can view $-(K+V)$ as the generator of some unknown process. Then, by theorem 2.8, this process forms a unique solution to the above differential equation. More specifically, the unique solution to the above differential equation follows from the semigroup $e^{-\tau(K+V)}$ corresponding to generator $-(K+V)$ and the initial condition $\rho(\mathbf{r},0) = \psi(\boldsymbol{r},0)$, yielding

$$\rho(\mathbf{r},t) = e^{-\tau(K+V)}\rho(\mathbf{r},0); \tag{9}$$

and thus we have solved the Schrödinger equation. Specifically, we have a formal solution to the Schrödinger equation in terms of the semigroup $e^{-\tau(K+V)}$ with generator $-(K+V)$.

6.2 Finding the ground-state

In order to find the ground-state, we will first consider the behaviour of eigenstates of equation 9. For an eigenstate ρ_E, we know that $(K+V)\rho_E = E\rho_E$ and so it follows that

$$e^{-\tau(K+V)}\rho_E = e^{-\tau E}\rho_E. \tag{10}$$

Now, note that the ground-state energy E_g is the smallest eigenvalue. Thus $e^{-\tau E_g}$ is the largest factor by which an eigenstate is multiplied. Thus by repeatedly applying $e^{-\tau(K+V)}\rho_E = e^{-\tau E}\rho_E$ and renormalising, we eventually find the ground-state ρ_{E_g}.

Moreover, if we multiply our semigroup by the operator $e^{\tau E_g}$, we essentially shift the potential such that for ρ_{E_g}, it holds that

$$e^{-\tau(K+V-E_g)}\rho_{E_g} = e^{-\tau(E_g-E_g)}\rho_{E_g} = \rho_{E_g}.$$

So, we might view ρ_{E_g} as an invariant distribution of the operator $e^{-\tau(K+V-E_g)}$. Now we can use this fact to find E_g. Note that we are talking about ρ as a distribution, not the quantum mechanical wave function $|\psi|^2$. To do so, we do need to assume that ρ is positive. However, for the ground-state of bosons, this is not a problem. For fermions, the ground-state wave function can be negative and the assumption of positivity is not valid. Although there are ways to work around this, we will not detail them here.

Let us start out with some trial energy E_T. Then, we find that

$$e^{-\tau(K+V-E_T)}\rho_{E_g} = e^{-\tau(E_g-E_T)}\rho_{E_g}.$$

Thus, if we start out with a normalised ρ_{E_g}, applying our modified operator destroys this normalisation with a factor $g(E_T) = e^{-\tau(E_g-E_T)}$. So in case $E_T > E_g$, $g(E_T) > 1$ and if $E_T < E_g$, $g(E_t) < 1$. So if we end up with ρ_{E_g} being too large after applying $e^{-\tau(K+V-E_T)}$, we have to decrease E_T. On the other hand if we end up with ρ_{E_g} being too small after applying $e^{-\tau(K+V-E_T)}$, we have to increase E_T.

This indicates a general procedure for finding the ground-state and the appropriate ground-state energy. We start out with a random target energy E_T and an initial distribution ρ. We then repeatedly apply the operator $e^{-\tau(K+V-E_T)}$ boosting the dominance of the ground-state ρ_{E_g} in ρ as shown above, so that ρ goes towards ρ_{E_g}. Simultaneously, we update E_T in such a way that the operator $e^{-\tau(K+V-E_T)}$ preserves the normalisation of ρ, so that as ρ goes to ρ_{E_g}, E_T follows to E_g.

6.3 Evolving the Schrödinger equation

The above section underlines the fundamental idea behind diffusion Monte Carlo. However, for a practical implementation of the algorithm, we need to compute $e^{-\tau(K+V)}$. And although the solution given in equation 9 is a formal solution, it cannot be evaluated directly. Nevertheless, we do know that the operator $e^{-\tau K}$ is the semigroup of standard Brownian motion, which we can compute quite easily. Thus, we would like to factor the exponential $e^{-\tau(K+V)}$, so we can apply the operators $e^{-\tau K}$ and $e^{-\tau V}$ individually. Now, an operator exponential can only be factored if they operators commute, but K and V do not commute. To circumvent this problem, we will use an approximation. By the Baker-Campbell-Haussdorff formula, we have that

$$e^{-\tau(K+V)} = e^{-\tau(V+K)} = e^{-\tau V}e^{-\tau K} + \mathcal{O}(\tau^2).$$

Thus if we use small enough time steps, we can neglect the terms associated with $\mathcal{O}(\tau^2)$. The operator $e^{-\tau V}e^{-\tau K}$ is something we can compute numerically using random walkers.

6.4 Implementation of the algorithm

In diffusion Monte Carlo, we will estimate the distribution ρ by means of random walkers. Suppose we start out with M walkers. The idea is then that the density of walkers corresponds to the density of ρ. Then for small time steps $\Delta\tau$, we can evolve this distribution by $e^{-\Delta\tau(K+V-E_T)} \approx e^{-\Delta\tau(V-E_T)}e^{-\Delta\tau K}$ with E_T the trial energy.

Firstly, we evolve a walker according to $e^{-\Delta\tau K}$, which corresponds to simulating $\Delta\tau$ time of standard Brownian motion. So since $B_{\Delta\tau} \sim \mathcal{N}(0, \Delta\tau)$, we have to move the walker by a random vector distributed according to $\mathcal{N}(0, \Delta\tau)$.

Thereafter, we could try to account for the operator $e^{-\Delta\tau(V-E_T)}$ by assigning the ith walker a weight of $w_i = 1$ and updating it on each iteration as $w_i := e^{-\Delta\tau(V-E_T)} \cdot w_i$, where we recall that V depends on the location of the walker. Then the weight signifies how important the walker is in calculating the distribution ρ. So overall the normalisation of ρ becomes $\sum w_i/M$ and based on this normalisation we could update E_T.

However, this approach contains a problem, a lot of walkers will quickly walk away to regions with a very small weight $e^{-\Delta\tau(V-E_T)}$ and so their contribution would vanish. In order to combat this, we replace the weights of the walkers by a branching process. In this process, *poor* walkers die and *good* walkers replicate. By doing this with the right probabilities, we will see that the density of poor walkers and good walkers will give a similar density as was done by the weighted walkers, but notably, a lot less walkers will be needed to get an accurate density.

For every walker, we will calculate $q = e^{-\Delta\tau(V-E_T)}$. Then if $q < 1$, the walker survive with probability q. If $q > 1$, the walker will give birth to $\lfloor q \rfloor$ walkers with probability q or $\lfloor q - 1 \rfloor$ walkers otherwise. By giving birth, we specifically mean that the original walker is duplicated. So as an example, suppose we have $q = 1.4$, then there is 40% chance of 1 additional walker and otherwise there are 0 additional walkers. We can actually encapsulate this process in one procedure. Let r be a sample drawn from the uniform distribution on $[0, 1]$ denoted $U(0, 1)$. Then we place $s = \lfloor q + r \rfloor$ walkers at the location of the original walker. Note that it is possible that $s = 0$. This corresponds to the walker not surviving.

The last thing to do is to update E_T. The following is a rule that works good in practice. Let E_0 denote the initial guess for the energy, α an adjustable parameter, M the current number of walkers and $\tilde{M}$ the initial number of walkers. Then we, take

$$E_T = E_0 + \alpha \ln\left(\frac{\tilde{M}}{M}\right).$$

In practice, E_0 might be very far off to start. In that case, we might want to run the algorithm several times updating E_0 after each time.

For an overview of the entire diffusion Monte Carlo procedure in pseudo code, see algorithm 1.

6.5 Harmonic oscillator

As an example, we apply diffusion Monte Carlo to the 3D harmonic oscillator with the potential $V(r) = \frac{1}{2}r^2$. Then the exact ground-state wave function is given by $\psi(r) = \frac{1}{(2\pi)^{3/2}}e^{-r^2/2}$ with $E_g = \frac{3}{2}$.

Algorithm 1 Diffusion Monte Carlo

```
Put walkers in random postions;
while Not finished do
    for all walkers do
        Let x be the location of the walker;
        Update x according to 𝒩(0, Δτ);
        Let q := exp(−Δτ(V(x) − E_T));
        Let r be a sample from U(0, 1);
        Replace the walker by ⌊q + r⌋ walkers;
    end for
end while
```

To test diffusion Monte Carlo, a simulation was done with 500 walkers and 50000 time steps, where we have taken $E_0 = 1.5$, $\Delta\tau = 0.01$ and $\alpha = 1$. The diffusion Monte Carlo finds that $E_g = 1.502 \pm 0.005$. In addition, figure 3 gives the probability distribution of finding a walker a distance r from the origin according to diffusion Monte Carlo. The density of the exact ground-state wave function has also been included in this plot as a dashed line. This is given by $\psi(r) \cdot 2\pi r^2$, which is the radial density of the harmonic oscillator. The extra factor $2\pi r^2$ comes from the fact that the spherical shells further from the origin represent a large part of the wave function in correspondence with the volume of a thin spherical shell $2\pi r^2 \, \mathrm{d}r$.

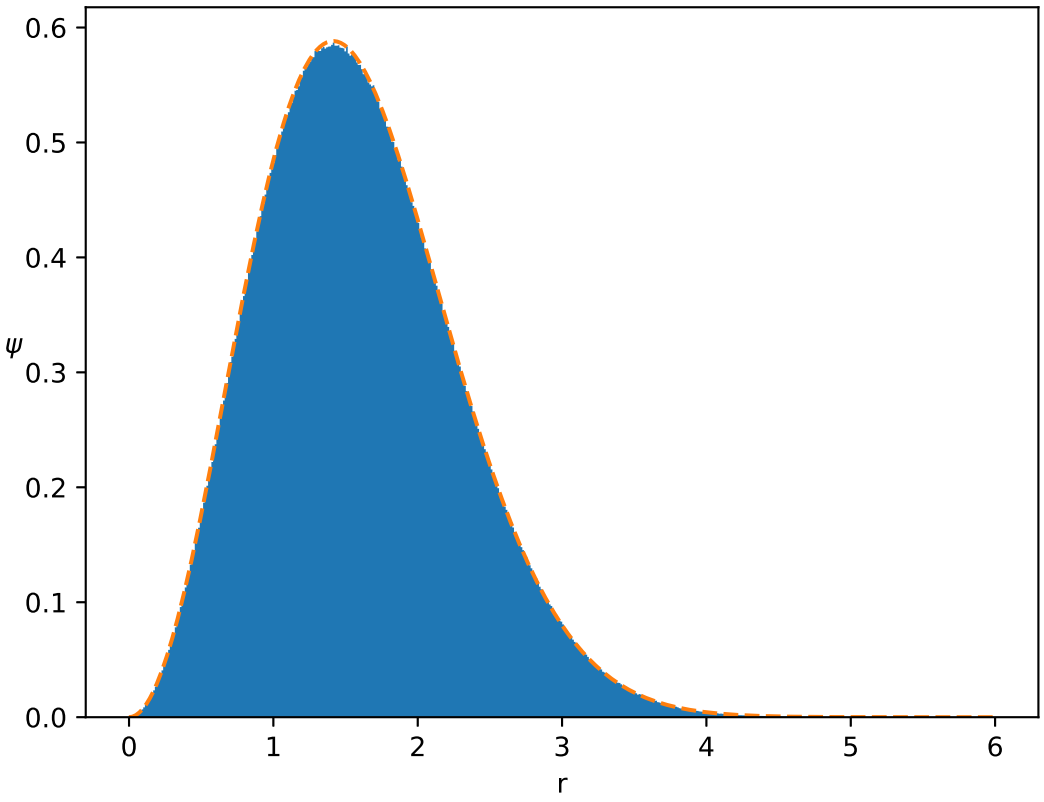

Figure 3: A plot of the density of the ground-state wave function for the potential $V(r) = \frac{1}{2}r^2$ as a function of the distance from the origin according to diffusion Monte Carlo in blue. The dashed orange line respresents the exact solution.

7 The Dirac Equation

During the 20th century, the Schrödinger equation proved to successfully predict a plethora of behaviours. However, apart from quantum mechanics there exists another very successful physical theory, namely special relativity. Now, the Schrödinger equation is not conform with special relativity. Thus, some alternative equation which does conform with special relativity was needed. The background of this chapter is based on [16]. Furthermore, the connection between stochastic processes and the Dirac equation was first shown in [17].

7.1 Relativistic Quantum Mechanics

One of the principle observation from special relativity is the energy momentum relation, namely

$$E^2 = p^2c^2 + m^2c^4, \tag{11}$$

where E is the particle's energy, p the particle's momentum, m the particle's mass and c the light speed. In quantum mechanics, we associate E with $i\hbar\frac{\partial}{\partial t}$ and p with $-i\hbar\nabla$. Thus, if we substitute the just mentioned operators into equation 11, we find the Klein-Gordon equation

$$-\hbar^2\frac{\partial^2\psi}{\partial t^2} = -\hbar^2c^2\nabla^2\psi + m^2c^4\psi.$$

Note that the above equation is not based on a first order time derivative, as the Schrödinger equation. It turns out that this leads to problems when describing particles with spin such as electrons. To solve this problem, another equation was introduced by Dirac. This equation starts from the principle that it should be based on a first order time derivative. From this, it follows that this equation should have the following form

$$i\hbar\frac{\partial\psi}{\partial t} = (-i\hbar\alpha\cdot\nabla + m\beta)\psi,$$

where α^i and β are to be determined and $\cdot$ denotes an inner product. Now, we still want that the above equation satisfies equation 11 and moreover, we want the equation to satisfy conservation of probability current. By this last statement, we mean that if we start out with a normalised wave function, it should stay normalised as we evolve the wave function according to the above equation.

It turns out these requirements cannot be satisfied if we take ψ to be $\mathbb{R}$ valued. However, if we consider it to be a 4-dimensional vector, we can satisfy these conditions. There are actually several options for α and β. Here, we choose to use the so called Weyl representation. Then the Dirac equations becomes as follows

$$i\hbar\frac{\partial\psi}{\partial t} = mc^2\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} + c\begin{pmatrix}\boldsymbol{\sigma} & 0\\ 0 & -\boldsymbol{\sigma}\end{pmatrix}\cdot\frac{\hbar}{i}\nabla\psi, \tag{12}$$

where $\boldsymbol{\sigma}$ represents a vector consisting of the Pauli matrices, so that $\boldsymbol{\sigma}\cdot\nabla = \sigma_x\frac{\partial}{\partial x} + \sigma_y\frac{\partial}{\partial y} + \sigma_z\frac{\partial}{\partial z}$. Recall that the Pauli matrices are given by

$$\sigma_x = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \quad \sigma_y = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \quad \sigma_z = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

The interpretation of the 4-components of the Dirac equation is a nuanced question. Generally, the Dirac equation is interpreted to represent a particle in superposition of two states. If we view the upper two components of the 4-dimensional vector ψ as a single particle, then the lower two components represent its anti-particle or a particle with opposite chirality. Now for both the upper and the lower particle, the two components associated with that particle represent the spin of the particle. In this sense, the Dirac equation is a very natural representation of an electron as it automatically results in the necessity of spin.

Now, the above equation represents the Dirac equation in 3-spatial dimensions. However, a generally easier case is to consider the one dimensional variant. This exists for the Dirac equation. However, in this case, we do not observe a spin, but only a chirality. So, we just have a 2-dimensional vector ψ. The 1-dimensional Dirac equation is as follows

$$i\hbar\frac{\partial\psi}{\partial t} = mc^2\sigma_x\psi - ic\hbar\sigma_z\frac{\partial\psi}{\partial x}. \tag{13}$$

7.2 Telegrapher process

Now, as with the Schrödinger equation, we want to connect the above equation to a stochastic process. This time, we will consider the stochastic process first and then work our way to the 1-dimensional Dirac equation.

Let us consider a free particle in 1-spatial dimension. This particle is subject to a stochastic process, where it propagates at a speed v and then at some rate a has a complete reversal of direction. Note that this exactly corresponds to the active particle discussed in example 5. Now, let $P_+(x,t)$ be the probability density of finding the particle moving to the right at location x and time t and $P_-(x,t)$ the probability density of finding the particle moving to the left at location x and time t. Then, for some infinitesimal time step Δt, we have

$$P_\pm(x, t+\Delta t) = P_\pm(x \mp \Delta x, t)(1 - a\Delta t) + P_\mp(x \pm \Delta x, t)a\Delta t.$$

So, we find that

$$\frac{P_\pm(x,t+\Delta t) - P_\pm(x,t)}{\Delta t} = \mp\frac{\Delta x}{\Delta t}\frac{P_\pm(x\mp\Delta x,t) - P_\pm(x,t)}{\mp\Delta x} + a(P_\mp(x\pm\Delta x,t) - P_\pm(x\mp\Delta x,t)).$$

Then, taking Δt and Δx to 0, we find

$$\frac{\partial P_\pm}{\partial t} = -a(P_\pm - P_\mp) \mp v\frac{\partial P_\pm}{\partial x}. \tag{14}$$

Remark 4. Another way to find the above equation is by letting L from example 5 act on distributions through its adjoint, which is similar to what happens in example 1. $\triangle$

If we now write out the components of equation 13 explicitly, we find that

$$i\hbar\frac{\partial\psi_\pm}{\partial t} = mc^2\psi_\mp \mp ic\hbar\frac{\partial\psi_\pm}{\partial x}.$$

Now, to connect the above equation with equation 14, we introduce the substitution $u(x,t) = e^{\frac{imc^2}{\hbar}t}\psi(x,t)$. Using this substitution, differentiation with $\frac{\partial}{\partial t}$ and immediately dividing out the phase factor $e^{\frac{imc^2}{\hbar}t}$, we find

$$mc^2 u_\pm + i\hbar\frac{\partial u_\pm}{\partial t} = mc^2 u_\mp \mp ic\hbar\frac{\partial u_\pm}{\partial x}.$$

From this, it directly follows that the $u_\pm$ satisfies

$$i\frac{\partial u_\pm}{\partial t} = -\frac{mc^2}{\hbar}(u_\pm - u_\mp) \mp ic\frac{\partial u_\pm}{\partial x}.$$

If we now transition, our equation to imaginary time $\tau = it$ and note that $c = \frac{\mathrm{d}x}{\mathrm{d}t}$, we find using $\frac{\partial}{\partial t} = \frac{\partial \tau}{\partial t}\frac{\partial}{\partial \tau} = i\frac{\partial}{\partial \tau}$ and multiplying the above equation by -1 that

$$\frac{\partial u_\pm}{\partial \tau} = -\frac{mc^2}{\hbar}(u_\pm - u_\mp) \mp c\frac{\partial u_\pm}{\partial x}, \tag{15}$$

which is exactly equation 14 with $v = c$ and $a = \frac{mc^2}{\hbar}$. So just like the Schrödinger equation is tied to Brownian motion, the Dirac equation is tied to an active particle process, where the reversal rates are determined by the mass of the particle. We will call A defined by $Au_\pm = \mp c\frac{\partial u_\pm}{\partial x}$ to be the propagator of the 1-dimensional Dirac equation.

We might even consider this idea, where we have a particle switching between two states and a propagator A, which dictates how a particles moves in each state, more generally. If we take $Au_\pm = \mp c\,\boldsymbol{\sigma}\cdot\nabla u_\pm$ with $\boldsymbol{\sigma}$ the vector of Pauli matrices as before, we have the propagator for the three dimensional Dirac equation. Note $u_\pm$ are both 2-dimensional vectors which represent spinors of a left-handed and right-handed version of the particle. The switching still occurs at the rate $\frac{mc^2}{\hbar}$. However, directly interpreting the propagation of the particle has been severely complicated as it is spatial movement have become tied to spin of the particle. In fact, it was an open question for quite a long time how exactly to interpret the 3-dimensional propagator, which was only entirely solved in 1999 [18].

Although interpretation of this equation may have been difficult, actually writing it down is not a challenge. So for the sake of completeness, the 3-dimensional imaginary time analogue of the Dirac equation is given by

$$\frac{\partial u_\pm}{\partial \tau} = -\frac{mc^2}{\hbar}(u_\pm - u_\mp) \mp c\,\boldsymbol{\sigma}\cdot\nabla\psi.$$

7.3 The Dirac equation with potential

One noteworthy thing is that up to now, we have only considered a free particle. One of the most interesting type of particles is a bound particle. So to study this we will introduce a potential $V(x,t)$. Let $\hat{K}$ denote $mc^2\sigma_x - ic\hbar\sigma_z\frac{\partial}{\partial x}$, which is the operator associated with free energy in the Dirac equation. We see that including a potential into Dirac equation comes down to $\hat{E}\psi = (\hat{K} + \hat{V})\psi$, which leads to

$$i\hbar\frac{\partial \psi_\pm}{\partial t} = mc^2\psi_\mp \mp ic\hbar\frac{\partial \psi_\pm}{\partial x} + V(x,t)\psi_\pm.$$

Then using the substitution $u(x,t) = e^{\frac{imc^2}{\hbar}t}\psi(x,t)$, transitioning to imaginary time and just repeating all the other steps to go from equation 13 to 15, we find that

$$\frac{\partial u_\pm}{\partial \tau} = -\frac{mc^2}{\hbar}(u_\pm - u_\mp) \mp c\frac{\partial u_\pm}{\partial x} - \frac{V(x,\tau)}{\hbar}u_\pm.$$

Now, use ϕ to denote $\frac{V}{\hbar}$. Then replacing $\frac{mc^2}{\hbar}$ by a, we find the following for the imaginary time analogue of the Dirac equation with potential:

$$\frac{\partial u_\pm}{\partial \tau} = -a(u_\pm - u_\mp) \mp c\frac{\partial u_\pm}{\partial x} - \phi(x,\tau)u_\pm. \tag{16}$$

In the next chapter, we will use techniques similar to those used in chapter 6 to construct a quantum Monte Carlo method for finding the ground state of the above equation.

8 A Monte Carlo Method for the Dirac Equation

In this chapter, we will combine the developed theory from the previous chapter with the idea of diffusion Monte Carlo from chapter 6 to find the ground-state of a 1-dimensional Dirac equation. For this, we will use equation 16 and quickly repeat the steps of the first three sections of chapter 6.

Firstly, we will assume that the potential $\phi(x,t)$ does not depend on time and write Φ for this operator. Moreover, we will write $Pu_\pm = a(u_\pm - u_\mp)$ for the state switching operator and $Ku_\pm = \pm c\frac{\partial u_\pm}{\partial x}$ for the linear propagator. Then, equation 16 becomes

$$\frac{\partial u_\pm}{\partial \tau} = -(P + K + \Phi)u_\pm.$$

Now, just as in chapter 6, solving the above equation corresponds to solving our original equation, the Dirac equation in this case. Moreover, we can view the operator $P+K-V$ as the generator of transition semigroup and thus by theorem 2.10, we can solve our differential equation using the semigroup $e^{-\tau(P+K+\Phi)}$ and the initial condition $u_\pm(x,0)$, yielding

$$u_\pm(x,\tau) = e^{\tau(P+K+\Phi)}u_\pm(x,0), \tag{17}$$

where we note that our semigroup is essentially a time dependent matrix acting on a vector containing the u_+ and u_- components.

Now, for an eigenstate $u_\pm^{\mathcal{E}}$ with eigenvalue $\mathcal{E}$, we find that $(P + K + \Phi)u_\pm^E = \mathcal{E}u_\pm^{\mathcal{E}}$ and so we have $(P + K + (\Phi - \mathcal{E}))u_\pm^E = 0$. From this, it immediately follows that the eigenvalue $\mathcal{E}$ is also an eigenvalue for the original Dirac equation, since we then have that

$$i\hbar\frac{\partial \psi_\pm}{\partial t} = mc^2\psi_\mp \mp ic\hbar\frac{\partial \psi_\pm}{\partial x} + (V(x,t) - \hbar\mathcal{E})\psi_\pm.$$

So, we see that the eigenvalue of the original Dirac equation is $E = \hbar\mathcal{E}$. From now on out, we will take $\hbar = 1$, so that $E = \mathcal{E}$. Now, similarly to chapter 6, we see that

$$e^{-\tau(P+K+\Phi)}u_\pm^E = e^{-\tau E}u_\pm^E.$$

This is exactly what we observed in equation 10. From this, it follows that we again observe that repeatedly applying the above operator amplifies the ground-state with eigenvalue E_g relative to the other eigenstates. If we now once again introduce the trial eigenenergy E_T, we find that

$$e^{-\tau(P+K+\Phi-E_T)}u_\pm^{E_g} = e^{-\tau(E_g-E_T)}u_\pm^{E_g};$$

and that the failure of the above equation to stay normalised indicates how close the trail energy is to the ground state energy, in exactly the same manner as was seen in chapter 6.

Thus, we can apply exactly the same algorithm as before except that the propagators of our walkers are different now. Using the Baker-Campbell-Haussdorff formula, we see that our propagator becomes

$$e^{-\tau(P+K+\Phi)} = e^{-\tau P}e^{-\tau K}e^{-\tau\Phi} + \mathcal{O}(\tau^2)$$

So once again if we use small time steps, we see that we can factor our propagation operator and apply the steps individually. We quickly recall that $e^{-\tau P}$ signifies a transition from the $+$ to the $-$ state at rate a, $e^{-\tau K}$ is a linear displacement in the states direction and $e^{-\tau\Phi}$ is just multiplication by a number.

8.1 Implementation of the algorithm

Again the idea is to estimate the distribution $u_\pm$ using random walkers. Assume we let the program start out with M walkers which are distributed randomly across both states and position. These walkers represent the density of $u_\pm$, where we recall that the individual components $u_\pm$ are not normalised, but the overall two component vector is. Then for small time steps $\Delta\tau$, we will evolve the walker distribution by $e^{-\tau(P+K+\Phi-E_T)} \approx e^{-\tau(\Phi-E_T)}e^{-\tau K}e^{-\tau P}$ with trial energy E_T.

Firstly, we consider the state transition according to P. Now, note that first order state transitions are $\mathcal{O}(\tau)$ and second order state transitions are $\mathcal{O}(\tau^2)$. Since we are neglecting, second order propagation anyways, it will not hurt to neglect second and higher order state transitions. Moreover, note that when considering a single state transition, we have a transition at rate a according to the exponential distribution during some small time step $\Delta\tau$. Thus, we can approximate this transition probability by $a\Delta\tau$.

Furthermore, to simulate the linear propagation, we simply displace in accordance with the operator $\pm c\frac{\partial}{\partial x}$; so we move a walker by $c\Delta t$ if it is in the right moving state and by $-c\Delta t$ if it is in the left moving state.

For the multiplicative operator $q = e^{-\tau(\Phi-E_T)}$, we once again introduce a branching process. So that we replace a walker by $s = \lfloor q + r \rfloor$ walkers, where r is a sample from $U(0,1)$.

Moreover, if we denote the current number of walkers by $\tilde{M}$ and start out at some energy E_0, we find that we take the target energy E_T to be

$$E_T = E_0 + \alpha \ln\left(\frac{\tilde{M}}{M}\right),$$

where α is an adjustable parameter.

Now, to simplify the procedure significantly, we will fix our units such that $\hbar = 1$ and $c = 1$. Thus, then the displacement operator reduces to moving by $\Delta\tau$ and the transition rate is equivalent to the mass m.

For an overview of the entire algorithm in pseudo code, see algorithm 2.

Algorithm 2 Dirac Monte Carlo

```
Put walkers in random positions and states;
while Not finished do
    for all walkers do
        Let x be the location of the walker and s its state;
        With probability mΔτ, s := −s;
        Update x := x + sΔτ;
        Let q := exp(−Δτ(V(x) − E_T));
        Let r be a sample from U(0, 1);
        Replace the walker by ⌊q + r⌋ walkers;
    end for
end while
```

8.2 Harmonic Oscillator

Consider the potential $\phi(x) = \frac{1}{2}\omega m x^2$. Now, we use our last degree of freedom in units to fix $\omega m = 1$, so that the potential of the system is given by

$$\phi(x) = \frac{1}{2}x^2.$$

Now, we observe that the only degree of freedom left in the system is the mass m, which is equivalent to state transition rate. We will consider two possibilities for m, namely $m = 1$ and $m = 0.1$. When $m = 1$, the particle will transition very often and thus be more diffusive, which is the classical limit. In addition, it is exactly the 1-dimensional case of the classical simulation from chapter 6 and we know it is the ground state energy $\frac{1}{2}$ in that case. On the other hand, when $m = 0.1$, the particle will travel a lot further before transitioning. This is the high energy behaviour and here we expect deviations from our typical ground-state function of the harmonic oscillator.

Now, in both simulation of 50000 time steps, the parameters $\alpha = 1$, $\Delta\tau = 0.001$ and $M = 5000$ were used. For $m = 1$, $E_0 = 0.5$ was used and it was found that $E_g = 0.48 \pm 0.007$ was found. Furthermore, figure 4 gives the probability density of finding a walker at location x. This conforms nicely to the classical solution, which has been plotted in a dashed orange line.

For $m = 0.1$, $E_0 = 1.1$ was used and it was found that $E_g = 1.09 \pm 0.02$. Moreover, figure 5 gives the probability density of finding a walker at location x. This distribution deviates strongly from the classical result plotted as a dashed orange line. It seems like the electrons are pushed more to side of the harmonic oscillator.

Further research might compare the above solutions to exact solutions of the Dirac equation or to the solutions from other numerical methods. Moreover, extending the considered numerical scheme to higher dimensions, especially 3-dimensions with propagator $e^{\pm\boldsymbol{\sigma}\boldsymbol{\nabla}}$, is of interest since Monte Carlo methods become more efficient compared to traditional numerical integration schemes for higher dimensions.

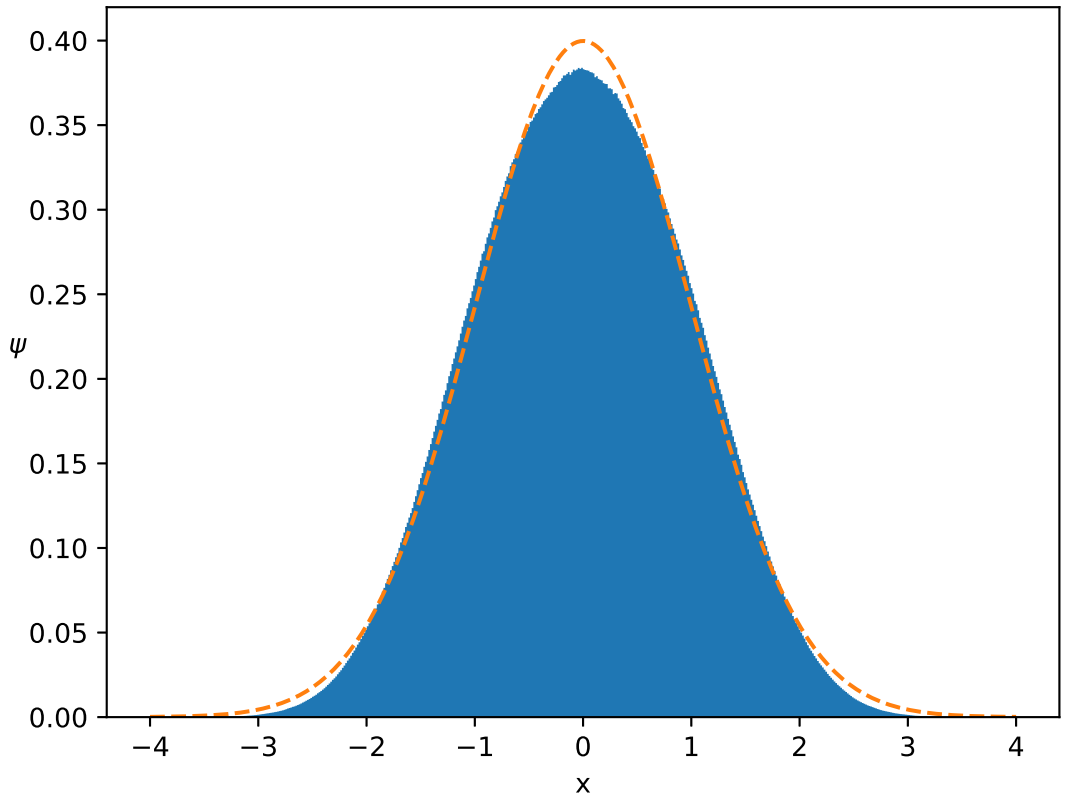

Figure 4: A plot of the density of the ground-state wave function of the Dirac equation for the potential $V(x) = \frac{1}{2}x^2$ and $m = 1$; produced by the Dirac Monte Carlo method in blue. The ground-state energy is given as $E_g = 0.48 \pm 0.007$. The dashed orange line represents the classical solution.

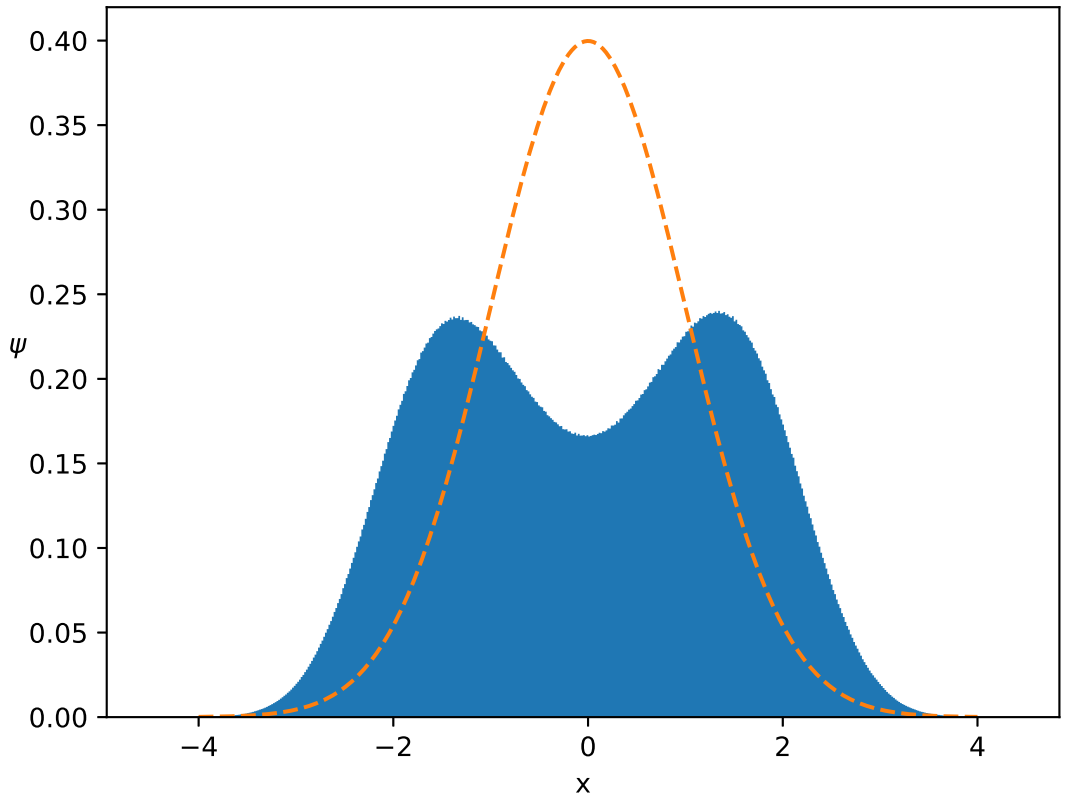

Figure 5: A plot of the density of the ground-state wave function of the Dirac equation for the potential $V(x) = \frac{1}{2}x^2$ and $m = 0.1$; produced by the Dirac Monte Carlo method in blue. The ground-state energy is given as $E_g = 1.09 \pm 0.02$. The dashed orange line represents the classical solution.

9 Conclusion

In this thesis, we developed a new technique based on homogenisation for calculating the diffusive limit of an active particle process. Firstly, this technique was used to find the diffusive limit of an $\mathbb{R}^d$-valued active particle process. The obtained result is in agreement with prior research, which found the same diffusive limit process. Thereafter, the on homogenisation based technique was used to obtain the diffusive limit of a manifold-valued active particle process. In this case, it is not clear, apriori, when the drift and diffusion of the obtained limit process are invariant under parallel transport i.e. constant. This matter requires further research.

In addition, a new numerical method for solving for the ground state of the 1-dimensional Dirac equation is demonstrated. This is a Monte Carlo method based on a link between the 1-dimensional Dirac equation and the telegrapher's process. It is a subject of further research to verify the results of this method by comparing the Monte Carlo method to an exact solution or other numerical methods. Furthermore, since Monte Carlo algorithms become relatively more efficient for problems of a high dimensionality, extending this algorithm to a higher dimensional Dirac equation is of interest.

References

[1] Bart van Ginkel, Bart van Gisbergen and Frank Redig. 'Run-and-tumble motion: the role of reversibility'. In: *Journal of Statistical Physics* 183.3 (2021), pp. 1–31.

[2] Erik Jørgensen. 'The central limit problem for geodesic random walks'. In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 32.1-2 (1975), pp. 1–64.

[3] R Versendaal. 'Large deviations for stochastic processes on Riemannian manifolds'. PhD thesis. Delft University of Technology, 2020.

[4] Isaac R Bruss and Sharon C Glotzer. 'Curvature-induced microswarming'. In: *Soft matter* 13.30 (2017), pp. 5117–5121.

[5] Leonardo Apaza and Mario Sandoval. 'Active matter on Riemannian manifolds'. In: *Soft matter* 14.48 (2018), pp. 9928–9936.

[6] Frank Redig. *Lecture notes on basic techniques in interacting particle systems*. 2014.

[7] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*. Springer, 2016.

[8] Olav Kallenberg. *Foundations of modern probability*. Vol. 2. Springer, 1997.

[9] Geoffrey Grimmett and Dominic Welsh. *Probability: an introduction*. Oxford University Press, 2014.

[10] Timo Seppäläinen. *Translation Invariant Exclusion Processes*. University of Wisconsin, 2008.

[11] Claude Kipnis and SR Srinivasa Varadhan. 'Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions'. In: *Communications in Mathematical Physics* 104.1 (1986), pp. 1–19.

[12] Tetsuro Nagai et al. 'Position-Dependent Diffusion Constant of Molecules in Heterogeneous Systems as Evaluated by the Local Mean Squared Displacement'. In: *Journal of Chemical Theory and Computation* 16.12 (2020). PMID: 33249840, pp. 7239–7254. DOI: 10.1021/acs.jctc.0c00448. eprint: https://doi.org/10.1021/acs.jctc.0c00448. URL: https://doi.org/10.1021/acs.jctc.0c00448.

[13] Philip Rosenau. 'Reaction and concentration dependent diffusion model'. In: *Physical review letters* 88.19 (2002), p. 194501.

[14] A. HINRICHS et al. 'THE CURSE OF DIMENSIONALITY FOR NUMERICAL INTEGRATION OF SMOOTH FUNCTIONS'. In: *Mathematics of Computation* 83.290 (2014), pp. 2853–2863. ISSN: 00255718, 10886842. URL: http://www.jstor.org/stable/24488680.

[15] J. Thijssen. *Computational Physics*. Cambridge University Press, 2007. ISBN: 9781139643764. URL: https://books.google.nl/books?id=xIkhAwAAQBAJ.

[16] J. Thijssen. *Lecture Notes Advanced Quantum Mechanics AP3051*. TU Delft, 2013.

[17] B. Gaveau et al. 'Relativistic Extension of the Analogy between Quantum Mechanics and Brownian Motion'. In: *Phys. Rev. Lett.* 53 (5 July 1984), pp. 419–422. DOI: 10.1103/PhysRevLett.53.419. URL: https://link.aps.org/doi/10.1103/PhysRevLett.53.419.

[18] B. Gaveau and L.S. Schulman. 'A Projector Path Integral for the Dirac Equation and the Spin Derivation of Space'. In: *Annals of Physics* 284.1 (2000), pp. 1–9. ISSN: 0003-4916. DOI: https://doi.org/10.1006/aphy.1999.5993. URL: https://www.sciencedirect.com/science/article/pii/S0003491699959936.

[19] Mark Veraar. *Lecture notes AM2090, Real Analysis*. Delft University Of Technology, 2020.

[20] Jean Jacod and Philip Protter. *Probability essentials*. Springer Science & Business Media, 2012.

A Measure Theoretic Probability

In order to fully grasp the contents of this thesis, some notions from probability theory are needed, which may not have been introduced to every mathematics bachelor's student. In this appendix, we will state some basic notions from measure theoretic probability without proof. The details can be found in the lecture notes for Real Analysis by Mark Veraar [19] and the first dozen chapters of Jacod's and Protter's book Probability Essentials [20].

In the following section, we will use Ω to denote an abstract space and 2^Ω to denote its powerset.

Definition A.1. *Then a collection $\mathscr{A} \subseteq 2^\Omega$ is called a* ***σ-algebra*** *if*

(i) $\emptyset, \Omega \in \mathscr{A}$

(ii) $A \in \mathscr{A} \implies A^c \in \mathscr{A}$

(iii) $(A_n)_{n\in\mathbb{N}} \subseteq \mathscr{A} \implies \bigcup_{n=0}^{\infty} A_n$

And for collection $\mathscr{F} \subseteq 2^\Omega$, we define the σ-algebra generated by $\mathscr{F}$ as

$$\sigma(\mathscr{F}) := \bigcap \{\mathscr{A} : \mathscr{A} \text{ is a } \sigma\text{-algebra and } \mathscr{F} \subseteq \mathscr{A}\}$$

From now on, we will consider $\mathscr{A}$ to be a σ-algebra on the space Ω; and for a topological space $(\Omega, \mathcal{O})$, we use the Borel σ-algebra denoted by $\mathcal{B}(\Omega) = \sigma(\mathcal{O})$ unless stated otherwise.

Definition A.2. *A function $\mu : \mathscr{A} \to [0, 1]$ is a* ***probability measure*** *if*

(i) $\mu(\Omega) = 1$

(ii) $\mu(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mu(A_n)$

A tuple $(\Omega, \mathscr{A}, \mu)$ is called a probability space.

In probability theory, the collection $\mathscr{A}$ is also called the event space. We then call $A \in \mathscr{A}$ an event or measurable set, as these are the sets from Ω for which μ defines a probability. In this light, we might also identify a σ-algebra with information, as we can only find a probability for its members. Furthermore, for an event $A \in \mathscr{A}$, we say it happens almost surely if $\mu(A) = 1$.

Definition A.3. *Suppose $(\Omega, \mathscr{A}, \mu)$ is a probability space and $(F, \mathscr{F}$ a measure spaces. Then a map $X : \Omega \to F$ is called a F-**valued random variable** if $X^{-1}(\mathscr{F}) = \{X^{-1}(A) : A \in \mathscr{F}\} \subseteq \mathscr{A}$.*

Thus random variables are functions on our space of outcomes Ω, which transfer the outcome from Ω to F. We then say that $\sigma(X) := X^{-1}(\mathscr{F})$ is the σ-algebra generated by the random variable X. That this is indeed a σ-algebra follows directly from the behaviour of unions and intersections under preimages. Intuitively, $\sigma(X)$ can be viewed as the information on the events in Ω through the outcomes of X.

It is now possible to transfer the probabilities from $(\Omega, \mathscr{A}, \mu)$ to $(F, \mathscr{F})$ through the law of X. This is the measure $\mu_X : \mathscr{F} \to [0, 1]$ defined by the map $A \mapsto \mu(X^{-1}(A))$. With the definition of a random variable out of the way, we can now define the expectation of a random variable using the lebesque integral.

Definition A.4. *The expectation of a random variable $X : \Omega \to \mathbb{R}^d$ is defined as*

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, \mu(\mathrm{d}\omega)$$

for a $\mathbb{R}$-valued random variable, we say it is integrable if $\mathbb{E}[|X|] < \infty$.

We say two measurable functions f, g are almost surely equal, if $\mu(\{\omega : f(\omega) \neq g(\omega)\}) = 0$. This is the probabilistic variant of almost everywhere equivalency. If $\mathcal{L}^p(\Omega, \mathscr{A}, \mu)$ with $p \geq 1$ is the space of measurable functions such that $\mathbb{E}[|X|^p] < \infty$ and $\sim$ denotes equivalence up to almost surety, we define $L^p(\Omega, \mathscr{A}, P) = \mathcal{L}^p(\Omega, \mathscr{A}, P)/\sim$.

We will now list some properties of expectations, which will be used throughout this thesis without further mention.

Proposition A.5. *Let $X, Y : \Omega \to \mathbb{R}^d$ be a random variables and $\alpha, \beta \in \mathbb{R}$.*

(i) $X \geq 0$ *a.s.* $\implies \mathbb{E}[X] \geq 0$

(ii) $X \geq 0$ *a.s. and* $\mathbb{E}[X] = 0 \implies X = 0$ *a.s.*

(iii) $\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$

(iv) *if f is a measurable function, then*

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)\, \mu_X(\mathrm{d}x)$$

(v) *if – in addition – X has a density $p_X : \mathbb{R}^d \to \mathbb{R}$, then*

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x) p_X(x)\, \lambda(\mathrm{d}x)$$

where λ denotes the Lebesque measure.

We can now introduce the concept of conditional expectation, which is used extensively in the theory of martingales and Markov processes.

Suppose we have thrown two dice under a table and our friend tells us that the total number of pips is 7. Then even though we do not know the exact configuration of our dice, we did obtain information about the outcome of our dice roll. Thus, we want to condition our random variable associated with the roll of two dice on the new information. To make this idea, rigorous, we define the conditional expectation.

Definition A.6. *Let $X : \Omega \to \mathbb{R}^d$ be an $L^1(\Omega, \mathscr{A}, \mu)$ random variable. Then for the sub-σ-algebra $\mathscr{F}$, we define the* ***conditional expectation*** *of X with respect to $\mathscr{F}$ as the unique $L^1(\Omega, \mathscr{F}, \mu|_{\mathscr{F}})$ random variable $\mathbb{E}[X|\mathscr{F}]$ such that*

$$\mathbb{E}[\mathbb{E}[X|\mathscr{F}]\mathbf{1}_F] = \mathbb{E}[X\mathbf{1}_F] \quad \forall F \in \mathscr{F}$$

where $\mathbf{1}_F$ is used to denote the indicator function of the set F.

Note that this definition is in actuality also a theorem, since we claim the conditional expectation exists and it is unique. Whereby unique, we mean unique in the μ-a.s. sense.

Proving the existence is beyond the scope of this thesis. However, it is possible either by using an object called the Radon-Nikodym derivative or by means of projections in L^2. For more information on this and conditional expectation in general, see the final chapters of Jacod's and Protter's Probability Essentials [20], which also form the basis of this section, or Kallenberg's Foundations of Modern Probability [8, pp. 163–184]. On the other hand, the uniqueness follows easily from the definition.

Proof of uniqueness. Suppose we have a $L^1(\mathscr{A})$ random variable X and two $L^1(\mathscr{F})$ random variables Y, Z satisfying definition A.6. Let A be the event $\{Y \leq Z\} \in \mathscr{F} \subseteq \mathscr{A}$. By definition, $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A]$. And thus we have that $\mathbb{E}[(Z-Y)\mathbf{1}_A] = 0$. Then as $(Z-Y)\mathbf{1}_A \geq 0$, we find $(Z-Y)\mathbf{1}_A = 0$ almost surely. Thus, we have $Y \leq Z$ almost surely. Similarly, we find $Y \geq Z$ almost surely. And thus $Y = Z$ almost surely. □

Although the conditional expectation is a random variable, its behaves in a sense like an expectation operator. The following proposition demonstrates this notion and aims to make the conditional expectation more intuitive.

Proposition A.7. *Let $X, Y \in L^1(\Omega, \mathscr{A}, \mu)$, $\alpha, \beta \in \mathbb{R}$ and $\mathscr{G} \subseteq \mathscr{F} \subseteq \mathscr{A}$ be a sub-σ-algebras. Then*

(i) $X \geq 0$ *a.s.* $\implies \mathbb{E}[X|\mathscr{F}] \geq 0$ *a.s.*

(ii) $\mathbb{E}[\alpha X + \beta Y|\mathscr{F}] = \alpha\mathbb{E}[X|\mathscr{F}] + \beta\mathbb{E}[Y|\mathscr{F}]$

(iii) $\mathscr{F} \perp\!\!\!\perp \sigma(X) \implies \mathbb{E}[X|\mathscr{F}] = \mathbb{E}[X]$

(iv) $\mathbb{E}[X|\{\emptyset, \Omega\}] = \mathbb{E}[X]$

(v) $\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[\mathbb{E}[X|\mathscr{F}]|\mathscr{G}]$

(vi) $\mathbb{E}[\mathbb{E}[X|\mathscr{F}]] = \mathbb{E}[X]$

(vii) *if Z is a $\mathscr{F}$ measurable random variable such that $XZ \in L^1$, then* $\mathbb{E}[XZ|\mathscr{F}] = Z\mathbb{E}[X|\mathscr{F}]$

Proof.

(i) Suppose $X \geq 0$ a.s. and consider $A = \{\mathbb{E}[X|\mathscr{F}] < 0\} \in \mathscr{F}$ and suppose $P(A) > 0$. Then $\mathbb{E}[\mathbf{1}_A] > 0$, so $0 > \mathbb{E}[\mathbf{1}_A\mathbb{E}[X|\mathscr{F}]] = \mathbb{E}[\mathbf{1}_A X]$. This is a contradiction, since $\mathbf{1}_A X \geq 0$ a.s. $\implies \mathbb{E}[\mathbf{1}_A X] \geq 0$. Thus $P(A) = 0$.

(ii) Let $Z = \alpha\mathbb{E}[X|\mathscr{F}] + \beta\mathbb{E}[Y|\mathscr{F}]$. Then for every $A \in \mathscr{F}$, we have $\mathbb{E}[\mathbf{1}_A Z] = \alpha\mathbb{E}[\mathbf{1}_A\mathbb{E}[X|\mathscr{F}]] + \beta\mathbb{E}[\mathbf{1}_A\mathbb{E}[Y|\mathscr{F}]] = \alpha\mathbb{E}[\mathbf{1}_A X] + \beta\mathbb{E}[\mathbf{1}_A Y] = \mathbb{E}[\mathbf{1}_A(\alpha X + \beta Y)]$. So $Z = \mathbb{E}[\alpha X + \beta Y|\mathscr{F}]$.

(iii) Suppose $\sigma(X) \perp\!\!\!\perp Y$ and let $A \in \mathscr{F}$. Then $\mathbb{E}[\mathbf{1}_A\mathbb{E}[X|\mathscr{F}]] = \mathbb{E}[\mathbf{1}_A X] \overset{\perp\!\!\!\perp}{=} \mathbb{E}[\mathbf{1}_A]\mathbb{E}[X]] = \mathbb{E}[\mathbf{1}_A\mathbb{E}[X]]$. Thus $\mathbb{E}[X|\mathscr{F}] = \mathbb{E}[X]$.

(iv) The σ-algebra $\{\emptyset, \Omega\}$ is independent from every sigma algebra. Thus by (iii), we have $\mathbb{E}[X|\{\emptyset, \Omega\}] = \mathbb{E}[X]$.

(v) Note that for all $A \in \mathscr{G} \subseteq \mathscr{F}$, it holds that $\mathbb{E}[\mathbf{1}_A \mathbb{E}[X|\mathscr{G}]] = \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[X|\mathscr{F}]] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[\mathbb{E}[X|\mathscr{F}]|\mathscr{G}]]$. Thus $\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[\mathbb{E}[X|\mathscr{F}]|\mathscr{G}]$.

(vi) Using (iv) twice and (v) for $\{\emptyset, \Omega\} \subseteq \mathscr{F}$, we find that $\mathbb{E}[X] = \mathbb{E}[X|\{\emptyset, \Omega\}] = \mathbb{E}[\mathbb{E}[X|\mathscr{F}]|\{\emptyset, \Omega\}] = \mathbb{E}[\mathbb{E}[X|\mathscr{F}]]$.

(vii) Suppose $Z = \mathbf{1}_B$ is $\mathscr{F}$ measurable. Then for every $A \in \mathscr{F}$, $\mathbb{E}[\mathbf{1}_A \mathbb{E}[\mathbf{1}_B X|\mathscr{F}]] = \mathbb{E}[\mathbf{1}_A \mathbf{1}_B X] = \mathbb{E}[\mathbf{1}_{A\cap B} \mathbb{E}[X|\mathscr{F}]] = \mathbb{E}[\mathbf{1}_A \mathbf{1}_B \mathbb{E}[X|\mathscr{F}]]$, since $B \in \mathscr{F}$. Thus also, by linearity, $\mathbb{E}[ZX|\mathscr{F}] = Z\mathbb{E}[X|\mathscr{F}]$ for all simple random variables Z.

Now, suppose $Z \geq 0$ is a random variable. Then there exists an increasing sequence Z_n of simple random variables such that $Z_n \uparrow Z$. In addition, assume $ZX \in L^1$. Then, we find that for all $A \in \mathscr{F}$, $\mathbb{E}[\mathbf{1}_A \mathbb{E}[ZX|\mathscr{F}]] = \mathbb{E}[\mathbf{1}_A ZX] \overset{MCT}{=} \lim \mathbb{E}[\mathbf{1}_A Z_n X] = \lim \mathbb{E}[\mathbf{1}_A Z_n \mathbb{E}[X|\mathscr{F}]] \overset{MCT}{=} \mathbb{E}[\mathbf{1}_A Z \mathbb{E}[X|\mathscr{F}]]$, where we used that we already know the identity holds for simple random variables. And thus $\mathbb{E}[ZX|\mathscr{F}] = Z\mathbb{E}[X|\mathscr{F}]$ for positive random variables Z.

Now, for Z an $\mathbb{R}$-valued random variable, the identity clearly holds by linearity after noting that $Z = Z^+ - Z^-$, where $Z^+, Z^- \geq 0$.

□

Example 9. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ be the outcome of a die throw and X be the random variable corresponding to this outcome. Now, let $\mathscr{F} = \sigma(\{\{1, 2\}, \{3, 4\}, \{5, 6\}\}) \subseteq 2^\Omega$. Then we must have that for $Y = \mathbb{E}[X|\mathscr{F}]$,

$$\begin{aligned}
\mathbb{E}[\mathbf{1}_{\{1,2\}} Y] &= \mathbb{E}[\mathbf{1}_{\{1,2\}} X] \\
\mathbb{E}[\mathbf{1}_{\{3,4\}} Y] &= \mathbb{E}[\mathbf{1}_{\{3,4\}} X] \\
\mathbb{E}[\mathbf{1}_{\{5,6\}} Y] &= \mathbb{E}[\mathbf{1}_{\{5,6\}} X]
\end{aligned}$$

Since Y is $\mathscr{F}$ measurable, this uniquely defines Y. $\mathbb{E}[X|\mathscr{F}]$ is X averaged over the smallest sets of the sub-σ-algebra. We can see this illustrated in figure 6. △

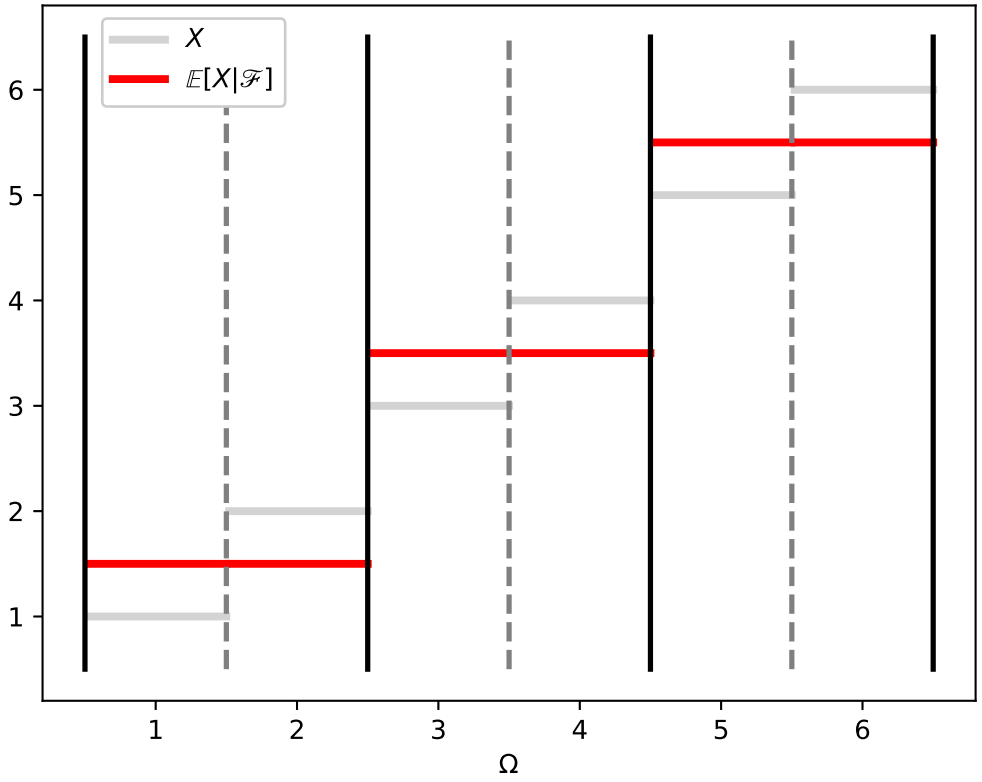

Figure 6: A plot showing the random variable $X : \{1, 2, 3, 4, 5, 6\} \to \mathbb{N}$ in grey and the random variable $\mathbb{E}[X|\mathcal{F}] : \{1, 2, 3, 4, 5, 6\} \to \mathbb{N}$ with $\mathcal{F} = \sigma(\{\{1, 2\}, \{3, 4\}, \{5, 6\}\})$ in red. The dashed grey lines delineate elements of the outcome space, the bold black lines delineate elements of $\mathcal{F}$.