

Delft University of Technology Faculty Electrical Engineering, Mathematics and Computer Science Delft Institute of Applied Mathematics

Invariance principle and hydrodynamic limits on Riemannian manifolds

A thesis submitted to the Delft Institute of Applied Mathematics in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE in APPLIED MATHEMATICS

by

Bart van Ginkel

Delft, the Netherlands June 2017

Copyright © 2017 by Bart van Ginkel. All rights reserved.



MSc THESIS APPLIED MATHEMATICS

"Invariance principle and hydrodynamic limits on Riemannian manifolds"

BART VAN GINKEL

Delft University of Technology

Daily Supervisor and responsible professor

Prof. Dr. F.H.J. Redig

Other thesis committee members

Prof. dr. J.M.A.M. van Neerven

Dr. W.M. Ruszel

June, 2017

Delft

Abstract

In this report we study Markov processes on compact and connected Riemannian manifolds. We define a random walk on such manifolds and give a direct proof of the invariance principle. This principle says that under some conditions on the jumping distributions (i.e. the distributions of single steps), the random walk converges to Brownian motion when space is scaled by $\frac{1}{N}$, time by N^2 and N tends to infinity (which has been shown with more general methods in [10] and [2]). To prove this, we show convergence of the generators on the set of smooth functions and we apply the Trotter-Kurtz theorem (as is done in [2], in a rather sketchy way and in a slightly different setting). We also show convergence of the corresponding Dirichlet forms. Then we show that the conditions on the jumping distributions are satisfied if they are compactly supported and have mean 0 and a covariance matrix which is invariant under orthogonal transformations.

Next, we define random grids on a Riemannian manifold and we define random walks on them. We show that their Dirichlet forms converge to the Dirichlet form of Brownian motion, using the results above. We also prove a result that is a bit weaker than convergence of the generators in this case.

Finally, these grids allow us to define the Symmetric Exclusion Process (SEP) on a manifold. Using the convergence results above, we follow a line of proof of [18] to show that the hydrodynamic limit of the SEP satisfies the heat equation. Some details still need to be filled in, but we believe that this method will allow us to study interacting particle systems and their hydrodynamic limits on Riemannian manifolds.

Before all of this we start with an introduction to Markov processes, their semigroups and generators. In particular we focus on time-reversible (or symmetric) processes and the Dirichlet form with its properties. We also give an introduction to Riemannian manifolds and related notions.

Preface

This report was written as the final thesis of the master Applied Mathematics at the TU Delft. It concludes five years of my life in which I have been introduced to mathematics by numerous book, courses, assignments and teachers. I must say I have gotten to appreciate mathematics, as will be readily confirmed by any one of my friends or family. I am very grateful for all the education that I have received in the past years. Multiple times during my studies, I have realized what a privilege it is to be able to study. To be in a situation where so much knowledge has been gathered over many years about all the mathematical topics that I am interested in and where all this knowledge is easily accessible through books, papers and professors. Already in my first year, I became interested in probability theory. After the course 'Advanced Probability' in my second year I was convinced that this was the department where I would want to graduate. Therefore my bachelor thesis was in this department (also supervised by Prof. Redig) and my master's program was mainly filled with probability courses. Since I like the (functional) analytic approach to problems in probability theory, it was a logical step to go to Prof. Redig again for my master thesis.

I embarked upon the project which gradually began to take shape. The first part consisted of studying two main topics: Dirichlet forms (and their connection to Markov processes and semigroups) and Riemannian geometry. Especially the latter was totally new to me, so I enjoyed discovering this field. I found out that it generalizes a lot of the concepts that were already familiar to me (such as differentiation and integration) in a way that made me understand them even better in the more familiar context. From then on, we started thinking about more direct proofs for already known results and new proofs for new results. In particular we wanted to find a way to define interacting particle systems on a manifold. I am happy that we have been able to formulate a way to do this, even though some parts of the proofs still need to be finished.

I would like to thank first of all Prof. Redig for your enthusiastic supervision and ideas throughout all of my project. As with my bachelor thesis, it was a good and instructive experience to be involved in your research. I would also like to thank my thesis committee for taking time to read my report. I hope that it will also be interesting for you. Further, I would like to thank my roommates at EWI for the fun and the nice conversations we had in between (and while) working on our projects. Thanks also to Rik for being available for questions and discussions. Last but not least, I would like to thank my friends, roommates, family and girlfriend for your patience and your interest whenever I was talking to you about random walks in curved spaces. I hope you will enjoy reading this report. I am more than happy to answer any questions about what I have written or to think with you about any new ideas related to this topic.

Bart van Ginkel

Delft, June 2017

Contents

Abstract 4							
Preface							
1	Introduction						
	1.1	The invariance principle	8				
	1.2	Hydrodynamic limits and grids	9				
	1.3	Introduction to background theory	11				
2	Markov processes and Dirichlet forms 13						
	2.1	Markov processes	13				
		2.1.1 Markov processes and semigroups	13				
		2.1.2 Symmetric Markov processes	18				
	2.2	Random walks and electrical networks	21				
		2.2.1 Potentials and hitting times	22				
		2.2.2 Energy dissipation	24				
	2.3	Dirichlet forms	26				
		2.3.1 Dirichlet forms	26				
		2.3.2 Correspondence with a semigroup	28				
		2.3.3 Meaning of the properties	29				
3	Introduction to Riemannian geometry 32						
	3.1	Riemannian manifolds	32				
		3.1.1 Inner product spaces	32				
		3.1.2 Riemannian metric	34				
		3.1.3 Integration	35				
		3.1.4 Geodesics	36				
		3.1.5 Exponential map	38				
		3.1.6 Normal coordinates	39				
	3.2	Connections and curvature	40				
	-	3.2.1 Connections	40				
		3.2.2 Curvature	44				
	33	Brownian motion	45				
	0.0	3.3.1 Laplace-Beltrami operator	46				
		3.3.2 Heat kernel	$\frac{10}{47}$				
		3.3.3 Brownian motion	48				
1	The	invariance principle on a Riemannian manifold	50				
т	1 1 1	Convergence of the generators	50				
	4.1	Convergence of the Dirichlet forms	57				
	4.2 4.3	Stepping distribution	57 60				
	1.0		00				
5	Ran	ndom grid	65				
	5.1	Model and motivation	65				
	5.2	Convergence of the Dirichlet forms	66				
		5.2.1 The first limit: a.s. convergence as $N \to \infty$	66				
		5.2.2 The second limit: $\epsilon \downarrow 0$	68				
		5.2.3 Conclusion and generalization	73				
	5.3	Convergence of the generators	77				

6	Hydrodynamic limit of the SEP					
	6.1	Symm	etric Exclusion Process	80		
	6.2	Hydro	dynamic limit	80		
	6.3	Conve	rgence result	83		
Summary and concluding remarks						
R	efere	nces		94		
\mathbf{A}_{j}	ppen	dix A	Introduction to differentiable manifolds	96		
	A.1	Differe	entiable manifolds	96		
		A.1.1	Smooth manifolds	96		
		A.1.2	Differentiation	98		
	A.2	Tanger	nt bundle	99		
		A.2.1	Intuitive definition using imbeddings	100		
		A.2.2	Vector bundles	101		
		A.2.3	Three constructions	101		
		A.2.4	Orientation	103		
	A.3	Vector	fields and differential equations	103		
	A.4	Differe	ential forms and integration	105		
		A.4.1	1-forms: intuitive introduction	105		
		A.4.2	The cotangent bundle and 1-forms	106		
		A.4.3	Covariant and contravariant tensor fields	107		
		A.4.4	Alternating functions and k-forms	109		
		A.4.5	Integration	111		

1 Introduction

In this report we study Markov processes on compact and connected Riemannian manifolds. In \mathbb{R}^n , there are a lot of well-known results about Markov processes. Two of them are the socalled invariance principle and the fact that hydrodynamic limits of certain interacting particle systems are given by the heat equation. In this report we deal with the question whether these results can be generalized to spaces that are only locally like \mathbb{R}^n , i.e. to manifolds. Section 2 and 3 and the appendix contain background material. Section 4, 5 and 6 contain our own research. In this section we will introduce the invariance principle and hydrodynamic limits in \mathbb{R}^n , motivate why they are relevant and speak about our aims and results in the context of a manifold. We will also give an overview of the content of all sections.

1.1 The invariance principle

The invariance principle in \mathbb{R}^n

Let μ be a distribution on \mathbb{R}^n with mean vector 0 and with the identity matrix as covariance matrix. We define a random walk X in \mathbb{R}^n as follows. The process starts at some initial point, say 0. Every time the process waits for an exponential time with parameter 1 and then jumps by adding a vector that is drawn from μ (all independently). The resulting process is a Markov process. Now suppose that for every $N \in \mathbb{N}$ we obtain a scaled process X^N as follows.

- Space is scaled by a factor $\frac{1}{N}$, i.e. at every jump we add $\frac{1}{N}$ times the random vector.
- Time is scaled by a factor N^2 , i.e. the rate of the exponential distributions is N^2 .

The invariance principle (or Donsker's theorem) states that, as N goes to infinity, the processes X^N converge in distribution in the path space to Brownian motion. In particular, X_t^N converges to the normal N(0,tI) distribution, where tI is t times the identity matrix. This is a version of the central limit theorem.

To understand intuitively why these scales are the right ones, consider the same situation in \mathbb{R} . Then μ has mean 0 and variance 1. Suppose that we scale space by a factor $\frac{1}{N}$. Then the jumps have variance $\frac{1}{N^2}$. Now scale time by a factor N^2 , so the exponential distributions have rate N^2 . Then in the time interval [0, 1] we expect around N^2 jumps. This means that the variance at time 1 is about $\frac{N^2}{N^2} = 1$. Since Brownian motion at time 1 has a N(0, 1) distribution, this is the variance that we want. These are heuristics, but they can be made precise. If we scale time by a larger factor than $O(N^2)$, the variance goes to infinity as N grows. If we scale time by a factor of $o(N^2)$, the variance goes to 0, so the limiting process will be identically 0.

Motivation

There are multiple reasons why such a result is important to obtain, also in the context of a manifold. First of all, this kind of approximation must be possible for any sensible definition of Brownian motion. In other words, if this does not hold, one should start wondering whether the definition of Brownian motion that is used is a good one. The reason is that the jumping process formalizes the idea of a combination of a lot of very small and independent movements, which is the main motivation and intuition behind Brownian motion (think of the famous pollen grains in the water of Robert Brown). We want Brownian motion to be the limit (in some sense) of such processes.

Further, such a result is necessary for simulation. If we simulate Brownian motion by using the jumping processes described above, we want to be sure that these indeed approximate Brownian motion in some sense.

The third reason is that Brownian motion is a very important and basic process in the sense that

a lot of other processes and partial differential equations are related to it or even constructed with it. Having an approximation result for Brownian motion can therefore be a handy tool for proving other results, such as the macro behaviour of interacting particle systems (we will see that the results of section 4 will be used for a result that is a bit different in section 5, which will be key to our main result in section 6).

Generalization to Riemannian manifolds

We can now ask whether the same result is true in some curved space M. First of all we need to properly define the processes that are involved. Since we need to walk 'straight' in a certain direction, we need to assume that M is a Riemannian manifold. Moreover we need to assume that it is compact and connected to avoid technical difficulties. All the necessary notions in this context are defined in section 3. In particular, section 3.3 describes how Brownian motion can be defined on a manifold. Now we associate to every point p of the manifold a distribution μ_p on the tangent space T_pM . We define a jumping process X as follows. Every time the process waits an exponential time with parameter 1, then picks a vector from T_pM according to μ_p and follows the geodesic in the direction of this vector for time 1 (again, all independently). Now we can again obtain scaled versions X^N as follows.

- Space is scaled by a factor $\frac{1}{N}$, i.e. at every jump we follow the geodesic for time $\frac{1}{N}$.
- Time is scaled by a factor N^2 , i.e. the rate of the exponential distributions is N^2 .

It has been shown in [10] and in [2] that under some restrictions on the jumping distributions we have the same result as in the Euclidean case: the processes X^N converge to Brownian motion in distribution in the path space. However, [10] uses a rather general method in a much more general context and [2] defines the process a bit differently (regarding both the current point and the next jump vector as the current state of the process) and uses a very specific jumping distribution (the uniform distribution on the unit vectors). Moreover, the proof in [2] is rather sketchy. We would like to understand more clearly why and under which conditions the convergence result holds. This is why section 4 is dedicated to give a direct proof in the situation that was sketched above.

Overview of section 4

We study the convergence from two different perspectives. In section 4.1 we show convergence of the generators of the processes and we use the Trotter-Kurtz theorem (as described in section 2.1). In section 4.2 we show convergence of the Dirichlet forms (in fact, of expressions that are proper Dirichlet forms when the underlying processes are symmetric). While following these proofs, we encounter several constraints that we need on the jumping distributions. In section 4.3 we discuss which distributions satisfy these constraints and provide examples. We show that a jumping distribution must be compactly supported and in local coordinates the mean vector must be 0 and the covariance matrix must be invariant under orthogonal transformations. In that case the processes converge to Brownian motion that is speeded by a constant. In particular the compactly supported measures that are invariant under orthogonal transformations have this property.

1.2 Hydrodynamic limits and grids

Hydrodynamic limit in \mathbb{R}^n

Next we turn to a problem that seems different, but will turn out to be very much related. First we consider \mathbb{R}^n again. Let $G = \mathbb{Z}^d$ be the grid consisting of the points with integer components. We can define the Symmetric Exclusion Process (SEP) on this grid as follows. First place some (possibly countably infinite) amount of particles on the grid in such a way that there cannot be more than one particle in a point. Then each particle jumps after independent exponential times (independent of the other particles) with the restriction that jumps to already occupied

points are cancelled. Let η_t denote the (random) configuration of particles at time t. Now we scale time and space again and obtain η^N for any $N \in \mathbb{N}$.

- Space is scaled by a factor $\frac{1}{N}$, i.e. we consider the grid $G^N = \frac{1}{N}G$.
- Time is scaled by a factor N^2 , i.e. the rate of the exponential distributions is N^2 .

Instead of looking at the amount of particles, we can study their (rescaled) mass distribution (in a way that will be made precise later). Suppose now that at time 0 this distribution converges to an absolutely continuous distribution $\rho_0 dx$ where $0 \leq \rho_0 \leq 1$. This situation is studied in [18, Chapter 8]. The author proves that under some restrictions on the jumping distribution (symmetry, translation invariance and bounded jumps), at every time point t the particle distribution converges to some absolutely continuous $\rho_t dx$. Moreover, $t \mapsto \rho_t$ solves the heat equation in \mathbb{R}^n . The limiting process is called the hydrodynamic limit.

Motivation

The main point of a hydrodynamic limit is that it allows us to move from a micro scale (individual particles) to a macro scale (densities). For example liquids and gasses, but also heat can in some way be represented by particles that are moving around (or oscillators exchanging energy). There are far too many particles to do calculations or simulations with all of them. Moreover, we are not always interested in individual particles, but rather in properties on a macro level. For instance the pressure of a gas or how quickly heat spreads through an object. For this purpose a well known tool is to take the limit of the amount of particles to infinity and treat it like a continuum, of which the behaviour is usually described by a partial differential equation. In this way we loose track of individual particles, but we obtain a (deterministic) description of macroscopic properties. In some sense, this is an infinite dimensional version of the law of large numbers.

To do all of this, we need a way to describe the behaviour of the particles and their interactions. First of all, we discretize space by using a grid. We will motivate that shortly. Then we model the particle motion with the Symmetric Exclusion Process. In the SEP all particles want to perform random walks. The interaction is given by the fact that there cannot be multiple particles at the same site (we call this exclusion). This models a repulsive force between particles. The SEP is interesting because it has interaction, but it is basic enough so that the methods of section 6 work.

Grid on a manifold

We would like to have the same result on a manifold. However, the first problem we bump on is how to define a suitable grid on which the particles have to move. In \mathbb{R}^n their is a very natural grid and a very natural way to make the grid finer. On a manifold, there does not seem to be a natural way to do this. One might consider not to use a grid at all and just make more and more particles move on the manifold itself. However, this leads to several technical difficulties, such as properly defining and showing that no piling-up of particles in a finite region occurs. Moreover, it would not make much sense to define the SEP in this way. Since the probability of jumping to an occupied spot would be 0 for every particle at every jump, the particles would be performing independent random walks and the interaction would be lost.

To understand better when a grid is considered a good grid, we need to take into account what we will use it for. We want to copy the proof from [18, Chapter 8]. We will see (in chapter 6) that a crucial step in this proof is that the generators of random walks on the grids converge to the Laplace-Beltrami operator (the generator of Brownian motion) and that the Dirichlet forms converge to the Dirichlet form of Brownian motion. Therefore this is the property that we look for when choosing a grid. For certain objects (especially for some nice surfaces) it is easy to imagine some sort of grid on them. One could think of using for instance triangulations (as in [17]). However, it is rather hard to make such a grid precise and to calculate what the correct jumping distributions are to converge to Brownian motion (and triangulations lead to a lot of technicalities). Moreover, we would like to have a general approach for every compact and connected Riemannian manifold.

Random grid

The approach we introduce here is to independently sample a sequence of points from the uniform distribution on the manifold (which is possible by compactness). We then define the grid G^N as the first N elements of the sequence (note that this grid is random). Then we define the jumping rate from a point p to a point q on the grid G^N as $k(d(p,q)/\epsilon)$, where $k : [0,\infty) \to [0,\infty)$ is a decreasing function that we will call a kernel and $\epsilon > 0$ is the bandwidth. This idea comes from papers like [19], where it shown that the graph Laplacian of those grids converges to the Laplace-Beltrami operator when first N goes to infinity and then ϵ goes to 0. However, the authors in [19] assume that the manifold is imbedded in a Euclidean space and they use the Euclidean distance, which we do not want.

Overview section 5

In section 5.1 we introduce and motivate the random grids and the way in which we define random walks on them. Then in section 5.2 we prove convergence of the Dirichlet form. Here we follow the idea of the proof in [3, Section 4], where the convergence of the same object is studied (under a different name). However, we do not use imbedding in a Euclidean space, which complicates the proof a lot. Then in section 5.3, we study convergence of the generators and we obtain a pointwise convergence result. Moreover, in section 5.2 and 5.3, we show that there exists a fixed sequence in M for which the convergence results hold. In fact, the set of sequences for which this is true has probability 1 under the product of uniform measures on M. In these proofs we use the results of section 4.

Overview section 6

Now that we have a grid to work with, we can try to follow the proof of [18]. In section 6.1, we describe the Symmetric Exclusion Process. Then in section 6.2, we introduce the initial definitions and assumptions and we explain in more detail that the convergence result we look for is weak convergence of random trajectories of measures. We also introduce the weak version of the heat equation, which must be satisfied by the hydrodynamic limit. The aim of section 6.3 is to complete the proof as is done in [18]. Some details are still to be completed, but the main line of the proof holds and is described in this section.

1.3 Introduction to background theory

Since we deal with a lot of concepts from the theory of Markov processes and corresponding operators and from Riemannian geometry, section 2 and 3 introduce these topics. The appendix contains a more general introduction to smooth manifolds.

Markov processes, generators and Dirichlet forms

We study Markov processes mainly through their associated generators and Dirichlet forms. From these two the generators are most commonly used. However, in some cases the Dirichlet form can be an easier object to work with. This is why we chose to also use this point of view. Section 2.1 is about Markov processes, semigroups, generators and in particular symmetric processes. In section 2.3 we introduce Dirichlet forms and their relation to Markov processes. To better understand this somewhat less natural concept, this section is preceded by section 2.2, in which Dirichlet forms are presented in the discrete context through their connections with electrical networks.

Smooth manifolds and Riemannian geometry

Section 3 contains an introduction to basic concepts from Riemannian geometry that we will

use later. In section 3.1 the Riemannian metric is introduced, along with geodesics and the exponential map. Then section 3.2 deals with connections and curvature. Since we will work with Brownian motion a lot, section 3.3 describes Brownian motion on a manifold and some related operators. In all of this we assume that the basic concepts from differential geometry (such as the tangent space, differential forms and integration on a manifold) are known. The appendix contains an introduction to these more general topics.

2 Markov processes and Dirichlet forms

In this report we will study stochastic processes via Dirichlet forms. In the following we will first recap some definitions and theorems involving Markov processes and their semigroups. We will zoom in on symmetric Markov processes, since these are the processes that correspond to Dirichlet forms. Then we will look into the elegant connection between random walks and electrical circuits in order to better understand the theory of Dirichlet forms and their relation to Markov processes, which will thereafter be presented.

2.1 Markov processes

We start with some basic theory about Markov processes and their semigroups. Then we will have a closer look at symmetric processes. These descriptions are based on [14, 16, 23].

2.1.1 Markov processes and semigroups

Markov processes

Let (S, \mathscr{A}) be a measure space and let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t \in T}, \mathbb{P})$ be a filtered probability space. Denote by $b\mathscr{A} = \{f : S \to \mathbb{R} | f \text{ bounded}, \mathscr{A}/\mathscr{B}\text{-measurable}\}$ the space of bounded, measurable, realvalued functions on S.

<u>Definition 2.1</u>. Recall that we call a stochastic process $X = \{X_t\}_{t \in T}$ on Ω that is adapted to the filtration $\{\mathscr{F}_t\}_{t \in T}$ and takes values in S a Markov process if the following holds:

$$\forall s, t \in T \text{ with } s < t \text{ and } \forall f \in b\mathscr{A} : \mathbb{E}[f(X_t)|\mathscr{F}_s] = \mathbb{E}[f(X_t)|X_s] \tag{1}$$

We will call S the state space of the process. This definition says that when estimating the future given the past, it suffices to know the state at the present time. This expresses that a Markov-process is memoryless: how the process behaves at a certain time only depends on the past through the state at that time.

Write \mathbb{P}_{μ} for the law of the process started from initial distribution μ and denote the corresponding expectation by \mathbb{E}_{μ} . In the case that $\mu = \delta_x$ for some $x \in S$ we will simply write \mathbb{P}_x and \mathbb{E}_x .

Strongly continuous semigroup

Now additionally assume that S is a compact metric space. Write C(S) for the continuous, real-valued functions on S. Also assume that $T = [0, \infty)$. For every $t \in T$ and $f \in C(S)$ define $S_t f : S \to \mathbb{R}$ by $S_t f(x) = \mathbb{E}_x[f(X_t)]$. If $S_t f \in C(S)$ for every $f \in C(S)$, S_t is an operator on C(S) and we call X a Feller process. We will assume this from now on. We will also assume that there exists a right continuous version of the process (i.e. the paths are almost surely right continuous) and that this is the version that we are working with. Moreover, we will assume that X is time homogeneous, which means that the process $\{X_{t+s}|X_t\}_{s\geq 0}$ has the same distribution as the process that starts from the distribution of X_t at time 0. Before we give some properties of the collection of operators $\{S_t, t \geq 0\}$ that we just defined, we need the following definition.

<u>Definition 2.2</u>. A collection of bounded, linear operators $\{T_t, t \ge 0\}$ on a Banach space X is called a *strongly continuous semigroup* if the following hold:

- $T_0 x = x$ $\forall x \in X$
- $T_s(T_t x) = T_{s+t} x$ $\forall s, t > 0, x \in X$
- $||T_t x x||_X \to 0$ as $t \downarrow 0$ $\forall x \in X$

This definition requires the semigroup property: first applying T_t and then T_s should be the same as directly using T_{s+t} . The strongly continuous part concerns the last requirement: the semigroup has to be right continuous at 0 with respect to the norm $|| \cdot ||_X$ (instead of for instance only pointwise in the case that X is a function space). An important implication is that the path $\{T_t x, t \ge 0\}$ is continuous everywhere for each x (this is easy to prove using the semigroup property).

Proposition 2.3. The collection of operators $\{S_t, t \ge 0\}$, as defined above, is a strongly continuous semigroup on $(C(S), || \cdot ||_{\infty})$. Moreover, it satisfies:

- $S_t \mathbb{1} = \mathbb{1}$ $\forall t \ge 0$
- $f \ge 0 \implies S_t f \ge 0$ $\forall t \ge 0, f \in C_b(S)$

Here 1 denotes the constant 1 function and the inequalities are meant pointwise. With these extra properties, we say it is a Markov semigroup.

Proof. The linearity is given by the linearity of expectation. Boundedness follows directly from

$$||S_t f||_{\infty} = ||\mathbb{E} \left[f(X_t) \right]||_{\infty} \le ||f||_{\infty}.$$
(2)

The property $S_0 f = f \ \forall f \in C(S)$ and the additional properties that make the semigroup a Markov semigroup are trivially satisfied. The semigroup property follows from:

$$\mathbb{E}_x[f(X_{t+s})] = \mathbb{E}_x[\mathbb{E}_x[f(X_{t+s})|\mathscr{F}_t]] = \mathbb{E}_x[\mathbb{E}_x[f(X_{t+s})|X_t]] = \mathbb{E}_x[S_sf(X_t)] = S_tS_sf(x).$$

Here we first used the tower property of conditional expectation, then the Markov property, then the time homogeneity and the definition of S_s and the last equation is just the definition of S_t .

Now only the strong continuity remains. Since, when starting from $x, X_0 = x$ a.s., it follows from right continuity of X that $X_t \to x$ a.s. as $t \downarrow 0$. Since f is continuous, $f(X_t) \to f(x)$. Since f is bounded, we see from the dominated convergence theorem that $\mathbb{E}_x f(X_t) \to f(x) = \mathbb{E}_x f(X_0)$ as $t \downarrow 0$, so $S_t f(x) \to f(x)$ as $t \downarrow 0$. This ensures pointwise right continuity in 0. To make this into $|| \cdot ||_{\infty}$ -convergence, a more advanced argument is needed. [14] directs the reader to section 1 of chapter IX of Yosida (1980).

Note that (2) in particular implies that S_t is a contraction on C(S).

<u>Remark 2.4</u>. It is possible to extend this theory to cases where S is only locally compact. However, it will in general not be possible to use the space of continuous function, so one can use for instance the space of bounded, continuous functions $C_b(S)$ or all continuous functions that vanish at infinity $C_0(S)$ ($f \in C_0(S)$ means that it is continuous and for every $\epsilon > 0$ there is a compact set $K \subset S$ such that $|f| < \epsilon$ on $S \setminus K$).

Sub-Markov semigroup

The properties above can also be relaxed a bit, this is the following definition.

<u>Definition 2.5</u>. Let $\{S_t, t \ge 0\}$ be a strongly continuous semigroup on C(S) such that for all $f \in C(S)$ and all $t \ge 0$:

 $0 \le f \le 1 \implies 0 \le S_t f \le 1.$

Then we call $\{S_t, t \ge 0\}$ a sub-Markov semigroup.

The idea here is that a Markov process might not have infinite lifetime. For instance if a particle runs off to infinity in finite time (by accelerating sufficiently fast) or if there is a positive killing rate (see example 2.11). This means that 'mass is lost' with a certain probability along the way. The corresponding semigroup is then

$$S_t f(x) = \mathbb{E}_x(f(X_t) \mathbb{1}_{\tau > t}),$$

where τ is the lifetime of X. Since $\mathbb{1}_{\tau>t} \in [0,1]$, we see that the semigroup stays positivity preserving, but even for positive constant functions it might be decreasing. Indeed, we have $S_t \mathbb{1} = \mathbb{E}_x(\mathbb{1}_{\tau>t}) = \mathbb{P}_x(\tau > t)$. These properties are captured by the definition. Of course, a Markov semigroup must also be a sub-Markov semigroup. This is indeed the case. Suppose $\{S_t, t \ge 0\}$ is a Markov semigroup. If $0 \le f \le 1$, then $S_t f \ge 0$ by the positivity preserving property. Since $\mathbb{1} - f \ge 0$ we see by the same property that $S_t(\mathbb{1} - f) \ge 0$, which implies that $S_t \mathbb{1} \ge S_t f$. In other words: $S_t f \le 1$. Further, it is easy to show that a sub-Markov semigroup with the additional property that it maps $\mathbb{1}$ to $\mathbb{1}$ is a Markov semigroup.

Correspondence process and semigroup

There is also a converse to 2.3: every Markov semigroup has a Markov process associated to it. This is theorem 1.5 in [14]:

<u>Theorem 2.6</u>. Suppose $\{S_t, t \ge 0\}$ is a Markov semigroup on C(X). Then there exists a unique Markov process $\{X_t, t \ge 0\}$ such that $\forall t \ge 0, x \in X, f \in C(X)$:

$$S_t f(x) = \mathbb{E}_x f(X_t).$$

We thus have a one-to-one correspondence between Markov semigroups and Markov processes.

Generators and the Hille-Yosida theorem

We will now add another operator to this relation.

<u>Definition 2.7</u>. Denote

$$D(L) = \{ f \in C(S) : \lim_{t \downarrow 0} \frac{S_t f - f}{t} \text{ exists in } C(S) \}.$$

Now define $L: D(L) \to C(S)$ by

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

We call L the generator of S and D(L) its domain.

The generator can be regarded as the derivative of the semigroup at 0. In this way it prescribes the behaviour of the semigroup in an infinitesimal time interval. Intuitively, knowing the generator of a Markov process should be enough to know the whole process. Indeed, at every time point, the process is memoryless and can be regarded to start again from that point. Since

$$Lf \approx \frac{S_{\Delta t}f - f}{\Delta t} \implies S_{\Delta t}f \approx f + \Delta t Lf$$
 (3)

for $\Delta t \ll 1$, it might be possible to approximate the semigroup by applying this formula involving only the generator over very small time intervals:

$$S_{t+\Delta t}f = S_{\Delta t}(S_t f) \approx S_t f + \Delta t L S_t f,$$

suggesting

$$S_{t+\Delta t} \approx S_t + \Delta t L S_t.$$

We see that S_t grows proportionally to itself times a 'factor' L, which leads to the (at this point formal) expression:

$$S_t = e^{Lt}$$

In this way the generator determines the semigroup and we have seen that the semigroup determines the process. These (rather heuristic) arguments turn out to arrive at the right conclusion. In fact, for a bounded generator L, one can simply define the exponential function by its series expansion

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

and check that is has the right properties. For unbounded operators this expression does not necessarily make sense, so a more intricate argument is needed. The Hille-Yosida theorem (theorem 2.9 in [14]) gives this argument and ensures us that there is a one-to-one correspondence between semigroups and their operators:

Theorem 2.8 (Hille-Yosida). The semigroup can be obtained from the following formula, which makes a one-to-one correspondence with definition 2.7. For $f \in C(S), t \ge 0$:

$$S_t f = \lim_{n \to \infty} (I - \frac{t}{n}L)^{-n} f.$$

Note that for real x, we see:

$$\lim_{n \to \infty} (1 - \frac{t}{n}x)^{-n} = \lim_{n \to \infty} \frac{1}{(1 + \frac{-tx}{n})^n} = \frac{1}{e^{-tx}} = e^{tx},$$

so this definition intuitively matches with what we want. We will not give a sketch of the proof here, but we will introduce an important key to it.

Resolvent

<u>Definition 2.9</u>. For Markov generator L and $\lambda > 0$ define the resolvent $R(\lambda, L) := (\lambda I - L)^{-1}$, where I is the identity operator.

Often, the I is left out and we write $(\lambda - L)^{-1}$. This expression can be shown to be well-defined for Markov generators and it is obviously directly determined by it. Now note that

$$\left(I - \frac{t}{n}L\right)^{-n} = \left(\frac{t}{n}\left(\frac{n}{t}I - L\right)\right)^{-n} = \left(\frac{t}{n}R\left(\frac{n}{t},L\right)\right)^{-n}$$

so according to the previous theorem, the semigroup is determined by the resolvent of the generator. Now, since the semigroup determines the generator and the generator determines the resolvent, we say that they are in a one-to-one-to-one correspondence and the resolvent captures all information of the generator and semigroup and hence of the underlying Markov process. We obtain the correspondence diagram from figure 1.



Figure 1: Correspondence of process, semigroup, generator and resolvent.

To understand why the resolvent contain so much information about the semigroup, we note that the equation

$$(\lambda - L)^{-1} = \int_0^\infty e^{-\lambda t} e^{tL} dt = \int_0^\infty e^{-\lambda t} S_t dt$$

is valid for $\lambda > 0$, so the resolvent can be thought of as the Laplace transform of the semigroup. Moreover, if S_t is a contraction, we obtain the following estimate:

$$||(\lambda - L)^{-1}||_{\infty} = \left| \left| \int_0^\infty e^{-\lambda t} S_t f dt \right| \right|_{\infty} \le \int_0^\infty e^{-\lambda t} ||S_t f||_{\infty} dt \le \int_0^\infty e^{-\lambda t} ||f||_{\infty} dt = \frac{1}{\lambda} ||f||_{\infty}.$$

This estimate is crucial in the proof of the Hille-Yosida theorem.

Maximum principle

Another key element of the proof is the following concept.

Proposition 2.10 (Maximum principle). Let L be a Markov generator and let $f \in D(L)$. Suppose $\overline{x \in S}$ is such that $f(x) \leq f(y)$ for all $y \in S$. Then $Lf(x) \geq 0$. This implies that if $(I - \lambda L)f = g$ for some $\lambda \geq 0$, we have

$$\min_{x} f(x) \ge \min_{x} g(x)$$

Proof. Note that the minimums exist, since S is compact and f is continuous. The first statement is true since $S_t f(x) = \mathbb{E}_x f(X_t) \ge \min_y f(y) = f(x)$ for all $t \ge 0$, so

$$Lf(x) = \lim_{t \downarrow 0} \frac{1}{t} (S_t f(x) - f(x)) \ge 0.$$

The second statement follows from the fact that for x such that $f(x) = \min_y f(y)$, we have $Lf(x) \ge 0$, so

$$\min_{y} g(y) \le g(x) = (I - \lambda L)f(x) = f(x) - \lambda f(x) \le f(x) = \min_{y} f(y).$$

Especially the second statement is an important characterisation of candidates for Markov generators.

Example: continuous time random walks

Now suppose that S is discrete (and hence finite, since we assume compactness) and the process is a random walk with independent exponential waiting times. In that case we can write the generator L as

$$Lf(x) = \sum_{y} \lambda P_{xy}(f(y) - f(x)),$$

where λ is the jump rate and P_{xy} is the probability that the process jumps from x to y. The proof of this is a straightforward calculation, where one conditions on the number of jumps that has occurred up to time t and uses the fact that the probability of more than one jump is $O(t^2)$. The domain of L is just the set of all functions on S (which is the same as C(S)).

Note that in general, speeding time by a constant c is equivalent to multiplying the generator (or, as we will see later, the Dirichlet form) with that constant. Indeed, denoting the semigroup that is generated by cL as S^c (for some generator L), we see

$$S_t^c = e^{t(cL)} = e^{(tc)L} = S_{tc}$$

where S is the semigroup generated by L. Evolving for time t under the action of S^c thus corresponds to evolving for time tc under the action of S. In this way it makes sense that the jumping rate λ appears as a factor in the generator of the random walk above.

If the space is not discrete the generator takes a similar form:

$$Lf(x) = \lambda \int (f(y) - f(x))\mu_x(\mathrm{d}y).$$

Here λ is still the jumping rate and for each $x \mu_x$ is a jumping (probability) distribution on the state space. The domain is still the set of all continuous functions.

Importance of the domain

The following example shows that the domain of a generator is not just a technicality, but that it really determines the behaviour of the process. In fact it shows how boundary conditions to a process can be captured be the choice of the domain.

Example 2.11 . Let X be a process taking values in $[0, \infty)$ with generator $L = \frac{1}{2} \frac{\partial^2}{\partial x^2}$ and let in this example $C^2 = C^2[0, \infty)$ denote the twice continuously differentiable functions on $[0, \infty)$. In each of the following cases, X is Brownian motion on $(0, \infty)$, but its behaviour at the boundary 0 depends heavily on the domain of L.

- If $D(L) = \{f \in C^2 : f(0) = 0\}$, then X is ended when it reaches 0 ('killing').
- If $D(L) = \{f \in C^2 : f'(0) = 0\}$, then X is reflected in 0 ('reflection').
- If $D(L) = \{f \in C^2 : f''(0) = 0\}$, then X stays in 0 forever once it reaches 0 ('holding').

Trotter-Kurtz

For some properties involving Markov processes it suffices to check things on a smaller set than the domain: a core.

<u>Definition 2.12</u>. For a generator L, a subset \mathscr{D} of D(L) is called a *core* if for any (f, Lf) in the graph of L, there are (f_n, Lf_n) with $f_n \in \mathscr{D}$ such that $(f_n, Lf_n) \to (f, Lf)$ as $n \to \infty$.

We can now formulate an important theorem that relates the convergence of generators to the convergence of the corresponding semigroups and processes.

Theorem 2.13 ([11, Thm 19.25], Trotter-Kurtz). Let $(X^n)_{n=1}^{\infty}$, X be Feller processes on a compact space S. Let $(L^n, D^n)_{n=1}^{\infty}$, L be their generators and $(T^n)_{n=1}^{\infty}$, T their Markov semigroups, respectively. Let \mathscr{D} be a core for L. Then the following are equivalent.

- For all $f \in \mathscr{D}$ there is a sequence $(f^n)_{n=1}^{\infty}$ such that $f^n \in D^n$, $f^n \to f$ and $L^n f^n \to L f$.
- $T_t^n f \to T_t f$ for every continuous f uniformly for t in compact sets.
- If $X_0^n \to X_0$ in distribution, then $X^n \to X$ in distribution in $D([0,\infty), S)$.

Here $D([0,\infty), S)$ denotes the set of all maps $[0,\infty) \to S$ that are right continuous and have left limits (so all cadlag maps). It is equipped with the Skorohod metric.

Note that if $\mathscr{D} \subset D^n$ for all n, we can take $f^n = f$ in the first statement. This means that for showing convergence in distribution in that case, it suffices to show convergence of the generators on a core.

2.1.2 Symmetric Markov processes

We will now study some definitions and results involving invariant measures and symmetry of a process, since we will need these when dealing with Dirichlet forms: it will turn out that they correspond to symmetric processes.

Invariant measures

We begin with a very important notion in the theory of Markov processes. In the following, when speaking of a measure on S we mean a finite, positive measure on (S, \mathcal{B}) where \mathcal{B} denotes the Borel σ -algebra.

<u>Definition 2.14</u>. We call a measure μ invariant or stationary for X if

$$\int S_t f d\mu = \int f d\mu \qquad \forall f \in C(S).$$

Note that integrals of continuous functions on a compact set with respect to a finite measure always exist. The intuition is that from any starting point on μ -average nothing happens to f when the process evolves. The next definition is even stronger.

Symmetric processes

<u>Definition 2.15</u>. Let $X = \{X_t, t \ge 0\}$ be a Markov process on S with semigroup $\{S_t, t \ge 0\}$. Let μ be a measure on S. We call X (and its semigroup) symmetric or reversible with respect to μ if for all $f, g \in C(S)$ and for all $t \ge 0$:

$$\int S_t f g \mathrm{d}\mu = \int f S_t g \mathrm{d}\mu.$$

We will call a process (and its semigroup) *symmetric* or *reversible* if there exists a measure with respect to which it is symmetric.

The idea behind symmetry is that the distribution of the process started in μ is the same when running forwards or backwards through time. We now have the following easy observation.

Proposition 2.16. If X is reversible with respect to μ , it is invariant with respect to μ .

Proof. Simply set g = 1 in the definition of reversibility and use the fact that $S_t 1 = 1$.

Relation with the generator

We can relate symmetry directly to the generator of a process. In fact it suffices to look at a core of the generator.

Proposition 2.17. Let L be the generator of X. Then X is symmetric with respect to μ if and only if there is a core $\mathscr{D} \subset D(L)$ such that:

$$\forall f, g \in \mathscr{D}: \qquad \int Lfg \mathrm{d}\mu = \int fLg \mathrm{d}\mu. \tag{4}$$

Discrete case: detailed balance

Recall that in the case of a random walk in a finite space, we can write the generator L as

$$Lf(x) = \sum_{y} \lambda P_{xy}(f(y) - f(x)),$$

where λ is the jump rate and P_{xy} is the probability that a particle at x jumps to y.

Proposition 2.18 (Detailed balance). In this discrete case X is symmetric with respect to μ if and only if the detailed balance condition holds:

$$\mu(x)P_{xy} = \mu(y)P_{yx} \qquad \forall x, y \in S$$

Proof. In this case S is finite, so $\mathscr{D}(L) = C(S) = \mathbb{R}^S$. Since the set V of functions for which (4) holds is a linear space, we see that X is symmetric if and only if V is a linear subspace of C(S) containing a core. From the definition of the core it becomes clear that it must be dense in C(S), so any set containing it must also be dense. But the only dense linear subspace is C(S) itself, so we see that X is symmetric if and only if V = C(S).

Now assume that X is reversible. Then for any two points $x_1, x_2 \in S$ we must have that (4)

holds for $f = \mathbb{1}_{x_1}$ and $g = \mathbb{1}_{x_2}$. Filling this in with the generator defined above yields the detailed balance condition for x_1 and x_2 . On the other hand, if the detailed balance condition holds, we see that (4) is true for any two indicator functions of points. Since any function $f \in C(S)$ can be written $f = \sum_x f(x)\mathbb{1}_x$, it follows from linearity of V that (4) is true for any two functions on S.

Semigroup on $L^p(S,\mu)$

When there exist an invariant measure μ , it is possible to extend the semigroup to a semigroup on $L^p(S,\mu)$ for any $p \ge 1$. To see this, we first show that S_t is also a contraction on C(S) with respect to $||\cdot||_p$, which denotes the norm on $L^p(S,\mu)$.

Proposition 2.19. Let $\{S_t, t \ge 0\}$ be the semigroup on C(S) corresponding to a Markov process $\overline{X} = \{X_t, t \ge 0\}$ with invariant measure μ . Then S_t is a contraction with respect to $|| \cdot ||_p$ for each $t \ge 0$.

Proof. Let $t \ge 0$ and $p \ge 1$. We see for any $x \in S$ and $f \in C(S)$ with Jensen's inequality that

$$|S_t f(x)|^p = |\mathbb{E}_x f(X_t)|^p \le \mathbb{E}_x \left(|f(X_t)|^p \right) = \left(S_t \left(|f|^p \right) \right) (x).$$

Now, using the invariance with respect to μ , we see:

$$||S_t f||_p^p = \int |S_t f|^p \mathrm{d}\mu \le \int S_t |f|^p \mathrm{d}\mu = \int |f|^p \mathrm{d}\mu = ||f||_p^p,$$

which shows that $||S_t f||_p \leq ||f||_p$, which is what we wanted to prove.

We are dealing with a compact set and a finite measure, so $C(S) \subset L^p(S,\mu)$ as a dense subset. Since S_t is a contraction on C(S) with respect to $|| \cdot ||_p$, we can extend it to all of $L^p(S,\mu)$ by standard arguments. This way $\{S_t, t \ge 0\}$ is a semigroup on $L^p(S,\mu)$. Call its generator L_{μ} . It can be shown that L_{μ} naturally extends L: it is the closure of L in $L^p(S,\mu)$.

On $L^p(S,\mu)$ we still have $S_t f = \mathbb{E} f(X_t)$, although as a μ -a.s. statement. We only defined $S_t f$ as an L^p limit of continuous functions so one has to check that this is actually true. This can be done using standard techniques and we will not do this here. In particular we have $S_t \mathbb{1}_A(x) = \mathbb{P}_x(\mathbb{1}_A(X_t)) = \mathbb{P}_x(X_t \in A).$

Semigroup on $L^2(S,\mu)$: self-adjoint generator

As a special case of the above, S_t can be defined on $L^2(S,\mu)$. This will turn out to be the natural extension when dealing with symmetric processes, since then the defining property generalizes to

$$(S_t f, g) = (f, S_t g) \qquad \forall f, g \in L^2(S, \mu), t \ge 0$$

where (\cdot, \cdot) denotes the inner product $(f, g) = \int fg d\mu$ on $L^2(S, \mu)$. Note that requiring the relation above for L^2 functions is equivalent to requiring it for continuous functions, because of the construction above. We also see that requiring symmetry with respect to a measure μ in fact makes S_t a self-adjoint operator on $L^2(S, \mu)$.

Because of proposition 2.17, we also see that in the symmetric situation

$$(Lf,g) = (f,Lg) \qquad \forall f,g \in D(L), t \ge 0$$

This means that L is a symmetric operator. It is in general not self-adjoint since the domain of its adjoint is larger. However, this symmetry relation also holds for L_{μ} and this operator is self-adjoint. We will from now on denote L_{μ} as L.

Non-positive generator

In the L^2 setting we can say a lot more about the semigroup and its generator. We will mention only what will be used later.

Proposition 2.20. Let X be a μ -symmetric process with semigroup $\{S_t, t \ge 0\}$ and generator L. Then L is a non-positive operator, i.e. $(Lf, f) \le 0$ for all $f \in D(L)$.

Proof. First let $f \in L^2(S, \mu)$ and $t \ge 0$. By the semigroup property, μ -symmetry and the fact that S_t is an L^2 -contraction we have:

$$(S_t f, f) = (S_{t/2} S_{t/2} f, f) = (S_{t/2} f, S_{t/2} f) = ||S_{t/2} f||^2 \le ||f||^2 = (f, f).$$

Now take $f \in D(L)$. We see for $h \ge 0$:

$$0 \ge \frac{(S_h f, f) - (f, f)}{h} = \left(\frac{S_h f - f}{h}, f\right).$$

Since $f \in D(L)$, we know that $\lim_{h \downarrow 0} \frac{S_h f - f}{h}$ exists in $(C(S), || \cdot ||_{\infty})$ and equals Lf, so also in $L^2(S, \mu)$ (μ is finite, so $||f||_p^p = \int |f|^p d\mu \leq \int ||f||_{\infty}^p d\mu = ||f||_{\infty}^p \mu(S)$). By continuity of (\cdot, f) , we see:

$$0 \ge \lim_{h \downarrow 0} \left(\frac{S_h f - f}{h}, f \right) = \left(\lim_{h \downarrow 0} \frac{S_h f - f}{h}, f \right) = (Lf, f).$$

We conclude that $(Lf, f) \leq 0$ for all $f \in D(L)$.

We showed above that $(S_t f, f) \leq (f, f)$. Note that in fact $t \mapsto (S_t f, f)$ is decreasing. If $0 \leq s \leq t$ we see

$$(S_tf,f) = (S_{s/2}S_{t-s}S_{s/2}f,f) = (S_{t-s}S_{s/2}f,S_{s/2}f) \le (S_{s/2}f,S_{s/2}f) = (S_sf,f).$$

Reducibility

A final notion that we introduce is the following.

<u>Definition 2.21</u>. A process is called *reducible* with respect to the invariant measure μ if there is a disjoint partition $S = A_1 \cup A_2$ such that both A_i have positive μ -measure and if for any i

$$S_t \mathbb{1}_{A_i} = \mathbb{1}_{A_i}$$

or, in other words, for any $x \in A_i$

$$\mathbb{P}_x(X_t \in A_i) = 1$$
 for all $t \ge 0$.

If the process is not reducible, we call it *irreducible*.

Intuitively, X is reducible if there are at least two μ -significant sets from which the process cannot escape.

2.2 Random walks and electrical networks

So far, we have seen multiple objects that describe a Markov process. In section 2.3, we will introduce yet another one: Dirichlet forms. To better understand what such a form is and where it comes from, we will dedicate a few moments to study the rather surprising but quite natural relation between random walks on graphs and electrical circuits. On the way, we will also introduce some notions from potential theory. This description is mostly based on [4, 6].

2.2.1 Potentials and hitting times

Electrical network

We start by defining an electrical network as a connected, undirected graph G = (V, E) where V is the set of vertices and E is the set of edges. We assume that there cannot be more than one edge between any two nodes. We define $R_{xy} \in (0, \infty]$ to be the resistance between $x, y \in V$ (or of the edge (xy)) and $c_{xy} \in [0, \infty)$ the corresponding conductance, so $c_{xy} = (R_{xy})^{-1}$. The case $R_{xy} = \infty$ and $c_{xy} = 0$ represents vertices that are not joined by an edge. Denote by I_{xy} the current from x to y. Note that the resistance and conductance are symmetric $(R_{xy} = R_{yx}$ and $c_{xy} = c_{yx}$), but the current is antisymmetric $(I_{xy} = -I_{yx})$. We define a potential V on the graph to be any function $v : G \to \mathbb{R}$ that maps a point x to its potential v_x . This all is just a mathematical description of an electrical circuit.

Random walk on a graph

On the other hand define a random walk on the graph as the Markov process corresponding to the generator

$$Lf(x) = \sum_{y \in G} \frac{c_{xy}}{\mu_x} (f(y) - f(x)),$$

where $\mu_x = \sum_y c_{xy}$ (we silently assume that these numbers are finite, one can for instance just assume that the graph is finite). The process jumps from a point x to y at rate proportional to the conductance of (xy). We denote by $P_{xy} = \frac{c_{xy}}{\mu_x}$ the jump probability. Note that for each x these sum to 1, so $\{P_{xy}, y \in G\}$ is a probability vector. In other words, since μ_x is a normalizing constant, the process jumps away from x at rate 1. Note that it can (almost surely) only jump along edges. Indeed if there is no edge between x and y, we have $c_{xy} = 0$, so the process jumps from x to y at rate 0.

We can regard μ a measure on G giving mass μ_x to each singleton set $\{x\}$. This way the process is reversible with respect to μ , as can be check through the detailed balance condition. Indeed for any $x, y \in G$:

$$\mu_x P_{xy} = \mu_x \frac{c_{xy}}{\mu_x} = c_{xy} = c_{yx} = \mu_y \frac{c_{yx}}{\mu_y} = \mu_y P_{yx}.$$

Now it is easy to see that μ is also invariant:

$$\mu_x = \mu_x \sum_y P_{xy} = \sum_y \mu_x P_{xy} = \sum_y \mu_y P_{yx}.$$

Correspondence

Given the jump probabilities, one can construct an electrical network that corresponds to it (using the relations above). Since the jump probabilities only depend on the fraction $\frac{c_{xy}}{\sum_y c_{xy}}$, they are invariant to scaling of the conductance, whereas the electric network obviously is not. This implies that there are many circuits that correspond to the same process. However, when an electric network corresponds to a finite measure μ , one can (by rescaling) obtain a canonical version of the network: the one that makes μ a probability measure.

We now have a set of parameters on a graph and two different interpretations. Intuitively, the connection might make sense naturally since a high conductance means that the current through an edge will be strong and it also means that the jump probability along the edge is high, so the process will probably jump along the edge a lot of times. We will show that the connection it not limited to intuition.

Calculating potentials

Now say we pick a point a in the circuit and give it potential 1 and we pick another point b and give it potential 0. We want to know what the potentials will be in the other points of the

graph. To calculate these potentials we need two important laws from the theory of electric circuits:

$$\begin{array}{lll} \textit{Ohm's law:} & I_{xy} = \frac{v_y - v_x}{R_{xy}} = c_{xy}(v_y - v_x) \\ \textit{Kirchhoff's first law:} & \sum_y I_{xy} = 0 \end{array}$$

The first one shows how the current through an edge depends on the conductance/resistance of the edge and the potentials of the adjacent nodes. The second dictates that the total current into a node must equal the current that goes out. Combining these two laws shows that

$$0 = \sum_{y} I_{xy} = \sum_{y} c_{xy}(v_y - v_x) = \sum_{y} c_{xy}v_y - v_x \sum_{y} c_{xy} = \sum_{y} c_{xy}v_y - v_x\mu_x,$$

which implies that

$$v_x = \sum_y \frac{c_{xy}}{\mu_x} v_y = \sum_y P_{xy} v_y.$$

The potential function v that we are looking for should thus satisfy:

$$v_{a} = 1$$

$$v_{b} = 0$$

$$v_{x} = \sum_{y} P_{xy}v_{y} \qquad \forall x \notin \{a, b\}$$
(5)

Before we further study this situation, we take a look at a probabilistic problem.

Calculating hitting probabilities

Consider again the points a and b in our graph. When starting the random walk (as defined above) from a point x, we would like to know the probability p_x that the process hits a before it hits b. To be able to write this down efficiently, define the following stopping times:

$$\tau_{a} = \inf\{t \ge 0 : X_{t} = a\} \tau_{b} = \inf\{t \ge 0 : X_{t} = b\}$$

where $\{X_t, t \ge 0\}$ denotes the process. Let \mathbb{P} be its underlying measure. We can now write the desired probability as $\mathbb{P}_x(\tau_a < \tau_b)$. The law of total probability shows (together with the memorylessness of the process) that for any x that is not a or b

$$p_x = \mathbb{P}_x(\tau_a < \tau_b) = \sum_y P_{xy} \mathbb{P}_y(\tau_a < \tau_b) = \sum_y P_{xy} p_y.$$

If x is a or b we see directly that p_x equals 1 resp 0. We conclude that the function p that we want should satisfy:

$$p_{a} = 1$$

$$p_{b} = 0$$

$$p_{x} = \sum_{y} P_{xy} p_{y} \qquad \forall x \notin \{a, b\}$$

Dirichlet problem

As we observe, the function p that we are looking for should satisfy exactly the same equations as v in the problem before. (5) is a special case of the so-called Dirichlet problem. This problem looks for a function that has certain fixed values on the boundary of a set (a socalled Dirichlet boundary condition) and solves a partial differential equation in the interior. In the discrete setting of a graph the boundary can be chosen more arbitrarily and the PDE simplifies to the third condition of (5): equations relating the value in a point to the values in the points surrounding it. We will call functions that satisfy these equations harmonic functions.

Existence and uniqueness

The first question is if we have uniqueness. To answer this, we use the following (more generally valid) principle:

Theorem 2.22 (Maximum principle). Every solution to (5) takes on its maximum value on the boundary.

We will not prove this principle here, but refer to [6] for the proof in a graph context. With this principle, it is not hard to show uniqueness.

Theorem 2.23 (Uniqueness). There is at most one solution to (5).

Proof. We give a sketch of the proof. Suppose u and v satisfy (5). Then w = u - v can easily be shown to satisfy (5) with the constraint that it is 0 in a and b. The maximum principle tells us that $w \leq 0$ everywhere. In the same way $-w \leq 0$ everywhere. We conclude that $w \equiv 0$, so u = v.

The second question is whether there exists a solution at all for this problem. This is not so easy to answer for the general Dirichlet problem, but in our graph case this is true. See [4], for instance, for some ways to calculate or approximate the solution.

Correspondence of potentials and expected values at hitting times

So far we have seen that we obtained the same problem in both interpretations of the graph. Since they have a unique solution, the solution to both problems is equal. This, surprisingly, means that we can calculate the probability of hitting a before b from x by simply giving a potential 1, b potential 0 and measuring the potential at x. Actually, this idea can be generalized. Let A be any subset of G and give each node $a \in A$ a potential f(a). Then we can, again, ask ourselves what the potential is going to be in the other nodes. So we want to know f for $x \notin A$. Now define the hitting time

$$\tau_A = \inf\{t \ge 0 : X_t \in A\}.$$

One can show in a way similar to above that $\mathbb{E}_x f(X_{\tau_A}) = f(x)$. Note that in the 0-1 case above, this reduces to

$$f(x) = \mathbb{E}_x f(X_\tau) = f(a) \mathbb{P}_x(X_\tau = a) + f(b) \mathbb{P}_x(X_\tau = b)$$

= $1 \cdot \mathbb{P}_x(X_\tau = a) + 0 \cdot \mathbb{P}_x(X_\tau = b) = \mathbb{P}_x(X_\tau = a) = \mathbb{P}_x(\tau_a < \tau_b),$

so we still have $f(x) = \mathbb{P}_x(\tau_a < \tau_b)$.

<u>Remark 2.24</u>. The theory that is sketched here on graphs can (under certain conditions) be extended to the general Dirichlet problem. This includes the approach using the maximum principle to obtain uniqueness. Also the last result holds, which shows the beautiful fact that the solution of a Dirichlet problem in any interior point can be obtained by letting a suitable process run from that point until it hits the boundary and averaging over the values at the hitting time. In the most basic situation the PDE is the Laplace equation, the process is Brownian motion and the solutions are called harmonic functions in the most basic sense.

2.2.2 Energy dissipation

Energy dissipation: a first Dirichlet form

Instead of only looking at the potential in the points, we can investigate the energy that is

dissipated in the circuit. When a current flows through a resistor, it will produce warmth. We will need another formula concerning electric circuits:

Joule's law: $P = U \cdot I$

Here P is power, so the amount of energy that is dissipated every second. U is the voltage, so the difference in potential between the points adjacent to the edge. This shows that the energy that is dissipated (per second, we will from now on silently assume that that is what we mean to say) depends on the potential. Since also $I = U \cdot c$, we obtain as an expression for the energy dissipation between x and y with potential p:

$$U_{xy}I_{xy} = U_{xy}U_{xy}c_{xy} = c_{xy}(p_y - p_x)^2.$$

We have this equation for every pair of nodes (note that non-connected nodes have conductance 0 between them, so no energy is dissipated), so we need to sum over this to get the total dissipated energy. For any potential function f we write:

$$\mathscr{D}(f) = \frac{1}{2} \sum_{x,y} c_{xy} (f(y) - f(x))^2.$$
(6)

We can generalize this a bit to an expression involving two functions:

$$\mathscr{D}(f,g) = \frac{1}{2} \sum_{x,y} c_{xy}(f(y) - f(x))(g(y) - g(x)).$$

Note that the factor $\frac{1}{2}$ in front appears to not count all edges twice. Obviously $\mathscr{D}(f, f) = \mathscr{D}(f)$ and we will use this notation interchangeably. This expression is a first example of a Dirichlet form, the exact definition will follow in the next section. Actually, if we add the condition that the Dirichlet form of a constant should be 0, it can be shown that any Dirichlet form in this context can be (uniquely) written in this way. It thus determines the conductances of the corresponding network.

Dirichlet's principle: minimizing the energy dissipation

The dissipated energy is always positive, since it is a sum of positive terms (actually strictly positive, unless the potential is constant). However, for different potentials this energy is generally different. It would be interesting to know what happens to the energy dissipation if we only fix the potential in certain points. The answer lies in the following very important theorem.

Theorem 2.25 (Dirichlet's principle). Let $A \subset G$ and for every $a \in A$ fix a potential p_a . The harmonic function f that satisfies these boundary values is the unique function that minimizes

$$\{D(f): f(a) = p_a \ \forall a \in A\}.$$
(7)

Proof. The proof is based on a variational method. If one writes any other function that satisfies the given boundary conditions as f + h for some h that is 0 on A, one can show that $\mathscr{D}(f,h) = 0$. From this we obtain $\mathscr{D}(f+h) = \mathscr{D}(f) + \mathscr{D}(h) + 2\mathscr{D}(f,h) = \mathscr{D}(f) + \mathscr{D}(h) > \mathscr{D}(f)$ unless $h \equiv 0$.

Heuristic meaning for electrical networks and random walks

Now let us take some time to think about what this means. First of all, nature apparently wants to minimize the amount of energy that it costs to obey its own laws. This is a very basic (and beautiful) observation on which several principles in the calculus of variations are based (for instance the theory concerning the hamiltonian or langrangian function).

To see what this says about potentials and stochastic processes, we have a closer look at (6). It

is a weighted average of the sum of the squared distances between the potentials in neighbouring points. Intuitively, this means that the solution to the Dirichlet problem minimizes the distances between the potentials in neighbouring points and if the edge between two nodes has a high conductance it has priority over edges with small conductance. This makes sense, since with the same difference in potential, more energy will be dissipated when the conductance is higher. In terms of stochastic processes it is harder to find an intuitive meaning of energy. However, when there is a high conductance between two points, the probability that a particle jumps between them is higher. As soon as it jumps from one place to another, the expected value when hitting the boundary will change into the expected value at the point that it jumps to. Since the probability was high, the expected value was probably already quite close. This also shows that the expected values in neighbouring points with a high conductance between them should be closer than when there is a low conductance.

We see that the link between a random walk and the related energy function is not as naturally clear as with for instance its generator. However, the energy function dictates the whole process (since, as remarked above, it determines the conductances) and the behaviour of the process can be obtained by minimizing it.

<u>Remark 2.26</u>. The energy function above is an example of Dirichlet's energy. This, more generally, is a measure of how wildly a function varies. This Dirichlet energy is related to the general Dirichlet problem in the sense that the solution to the latter minimizes the former. In the most basic situation when dealing with the Laplace operator, the Dirichlet energy is $\mathscr{D}(f) = \frac{1}{2} \int (\nabla f)^2 d\lambda$. This will turn out be the Dirichlet form corresponding to Brownian motion, which might not be so surprising since remark 2.24 already showed the relation between the Laplace operator and Brownian motion.

2.3 Dirichlet forms

Now that we have seen a lot of the intuition behind it, we study the theory of Dirichlet forms. This will mostly be based on [1, Chapter 4] although at times we will use [5]. We will first give the necessary definitions, then we explain how a Dirichlet form corresponds to a semigroup and finally we will study the given definitions more closely in the context of the semigroup.

2.3.1 Dirichlet forms

Properties of the energy dissipation functional

Let us first have a closer look at the energy expression that we found above. Remember that this is the general expression for Dirichlet forms (that map constants to 0) in the finite context, which makes it more interesting than just a specific example.

$$\mathscr{D}(f,g) = \frac{1}{2} \sum_{x,y} c_{xy} (f(y) - f(x))(g(y) - g(x)).$$

We have already noted and used that $\mathscr{D}(f) = \mathscr{D}(f, f) \geq 0$ for any potential f. It is also clear that \mathscr{D} is symmetric: $\mathscr{D}(f,g) = \mathscr{D}(g,f)$. Finally note the following, less intuitive property. Suppose we take any f and define $g = 1 \land (f \lor 0)$ (so every value smaller than 0 is replaced by 0 and every value larger than 1 is replaced by 1). Let $x, y \in G$. If f(x), f(y) <= 0, we see $|g(y) - g(x)| = |0 - 0| = 0 \leq |f(y) - f(x)|$. Similarly, if f(x), f(y) > 1 we have $|g(y) - g(x)| \leq |f(y) - f(x)|$. In any other situation: if one of the values is larger than 1, the other is not, so replacing it by 1 brings them closer together. An analogous argument holds if one of them is smaller than 0. If they are both between 0 and 1 nothing changes. We conclude that in any situation $|g(y) - g(x)| \leq |f(y) - f(x)|$. Since all c_{xy} are positive, this directly shows that $\mathscr{D}(g) \leq \mathscr{D}(f)$. We will come back to this later, but the three properties that we mentioned here, suggest the following definitions.

Symmetric forms

Let S be a compact, separable metric space and let μ be a positive Radon measure on S with full support (i.e. $\operatorname{supp}(\mu) = S$). Recall that a Radon measure is a Borel measure that is inner regular and locally finite. Denote $H = L^2(S, \mu)$.

<u>Definition 2.27</u>. Let $\mathcal{E} : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ be a function where \mathcal{D} is some dense, linear subspace of H. We call $(\mathcal{E}, \mathcal{D})$ a symmetric form on H if

- $\mathcal{E}(f,g) = \mathcal{E}(g,f)$ $\forall f,g \in \mathcal{D}$
- $\mathcal{E}(f, f) \ge 0$ $\forall f \in \mathcal{D}$
- \mathcal{E} is linear in each coordinate

We call $\mathcal{D}(\mathcal{E}) := \mathcal{D}$ the domain of \mathcal{E} . Often we will write $\mathcal{E}(f) := \mathcal{E}(f, f)$ for the quadratic form induced by \mathcal{E} .

We can now for $\alpha > 0$ define \mathcal{E}_{α} on the same domain by $\mathcal{E}_{\alpha}(f,g) = \mathcal{E}(f,g) + \alpha(f,g)$, where $(f,g) = \int fg d\mu$ is the inner product on H.

<u>Proposition 2.28</u>. \mathcal{E}_{α} is again a symmetric form on H and it defines a norm on \mathcal{D} : $||f||_{\mathcal{E}_{\alpha}} = \mathcal{E}_{\alpha}(f)^{\frac{1}{2}}$. The metrics induced by \mathcal{E}_{α} and \mathcal{E}_{β} in this way are equivalent for $\alpha, \beta > 0$.

<u>Remark 2.29</u>. The definitions above can be generalized to arbitrary Hilbert spaces, but we will not need that throughout this report.

We now have the next set of definitions.

<u>Definition 2.30</u>. Let $(\mathcal{E}, \mathcal{D})$ be a symmetric form on *H*. We call it

- closed if $(\mathcal{D}, || \cdot ||_{\mathcal{E}_1})$ is complete
- Markov if for any $f \in \mathcal{D}$, $g := 1 \land (f \lor 0) \in \mathcal{D}$ and $\mathcal{E}(g) \leq \mathcal{E}(f)$

Dirichlet forms

We are now able to define a Dirichlet form.

<u>Definition 2.31</u>. We define a *Dirichlet form* to be a closed, Markov, symmetric form $(\mathcal{E}, \mathcal{D})$. Moreover we say it is

- regular if $\mathcal{D} \cap C_0(S)$ is dense in $(\mathcal{D}, \mathcal{E}_1)$ and in $(C_0(S), || \cdot ||_{\infty})$
- local if $\mathcal{E}(f,g) = 0$ whenever f, g have disjoint supports
- conservative if $\mathbb{1} \in \mathcal{D}$ and $\mathcal{E}(\mathbb{1}) = 0$
- *irreducible* if it is conservative and $\mathcal{E}(f) = 0$ implies f is constant

These are a lot of definitions. We will explain now how Dirichlet forms relate to Markov processes via semigroups and then we will try to provide some understanding or intuition behind the definitions above.

2.3.2 Correspondence with a semigroup

As the coming two theorems will describe, there is a one-to-one relation between Dirichlet forms and a certain class of processes, namely the symmetric sub-Markov processes. These theorems and part of the explanation around them are essentially [1, Theorem 4.5 and 4.6] and part of the surrounding text.

Dirichlet form corresponding to a semigroup

We first show how to obtain a Dirichlet form from a symmetric semigroup. Note that such a semigroup is a contraction semigroup, as was shown before.

Theorem 2.32 ([1, Thm 4.5]). Let $\{S_t, t \ge 0\}$ be a μ -symmetric sub-Markov semigroup on $L^2(S, \mu)$. Define for $f \in L^2(S, \mu)$ and t > 0:

$$\varphi_f(t) = \frac{(f - S_t f, f)}{t}.$$

Let

$$\mathcal{D} = \{ f \in L^2(S, \mu) : \lim_{t \downarrow 0} \varphi_f(t) < \infty \}$$
$$\mathcal{E}(f) := \mathcal{E}(f, f) := \lim_{t \downarrow 0} \varphi_f(t) \quad for \ f \in \mathcal{D}$$

Then $(\mathcal{E}, \mathcal{D})$ is a Dirichlet form. Moreover, if (L, D(L)) is the generator of $\{S_t, t \ge 0\}$, then $D(L) \subset \mathcal{D}, \mathcal{D}$ is dense in $L^2(S, \mu)$ and for $f \in D(L)$ and $g \in \mathcal{D}$:

$$\mathcal{E}(f,g) = (-Lf,g),\tag{8}$$

where $(f,g) = \int fg d\mu$ denotes the inner product on $L^2(S,\mu)$.

Since \mathcal{D} is dense in $L^2(S,\mu)$ we can in fact say that $(\mathcal{E},\mathcal{D})$ is a Dirichlet form on $L^2(S,\mu)$. Note that it suffices to only have a defining expression for $\mathcal{E}(f,f)$. Because of the relation

$$\mathcal{E}(f+g, f+g) = \mathcal{E}(f, f) + 2\mathcal{E}(f, g) + \mathcal{E}(g, g),$$

which follows from linearity and symmetry of \mathcal{E} , we can define $\mathcal{E}(f,g)$ using the polarization identity:

$$\mathcal{E}(f,g) = \frac{1}{2}(\mathcal{E}(f+g,f+g) - \mathcal{E}(f,f) - \mathcal{E}(g,g)).$$

It is also easy to see how (8) and the definition of \mathcal{E} are related:

$$(-Lf, f) = (-\lim_{t \downarrow 0} \frac{S_t f - f}{t}, f) = (\lim_{t \downarrow 0} \frac{f - S_t f}{t}, f) = \lim_{t \downarrow 0} \frac{(f - S_t f, f)}{t} = \lim_{t \downarrow 0} \varphi_f(t) = \mathcal{E}(f, f).$$

Semigroup corresponding to a Dirichlet form

Conversely, to each Dirichlet form corresponds a symmetric semigroup. This is what the next theorem says.

Theorem 2.33 ([1, Thm 4.6]). Let $(\mathcal{E}, \mathcal{D})$ be a Dirichlet form on $L^2(S, \mu)$. Then there exists a μ -symmetric sub-Markov semigroup $\{S_t, t \geq 0\}$ on $L^2(S, \mu)$ with generator (L, D(L)) and resolvent $\{U_{\lambda} = R(L, \lambda), \lambda > 0\}$ such that (8) is satisfied and for all $f \in L^2(S, \mu)$ and g in \mathcal{D} and for any $\alpha > 0$ we have

$$\mathcal{E}(U_{\alpha}f,g) + \alpha(U_{\alpha}f,g) = (f,g)$$

Proof. We give only a very brief sketch to obtain some idea of the proof. Using (8) and the definition of the resolvent we would like to have

$$(f,g) = ((\alpha - L)U_{\alpha}f,g) = (\alpha U_{\alpha}f,g) + ((-L)(U_{\alpha}f),g) = \alpha (U_{\alpha}f,g) + \mathcal{E}(U_{\alpha}f,g) = \mathcal{E}_{\alpha}(U_{\alpha}f,g)$$

Now, given \mathcal{E} (which is a closed, symmetric form), one can for each f define $U_{\alpha}f$ by this relation and the Riesz representation theorem and check that it satisfies the properties of a resolvent. From this one can deduce that there is a corresponding semigroup (as in theorem 2.8).

Relation with the generator

These two theorems show a one-to-one correspondence between Dirichlet forms on $L^2(S,\mu)$ and μ -symmetric sub-Markov semigroups on that same space. Theorem (2.33) only gives an implicit description of U_{α} . However, (8) provides us with a very useful and direct relation between a generator and its corresponding Dirichlet form. Returning to the case where the process is Brownian motion and the Dirichlet form is $\frac{1}{2} \int (\nabla f)^2 d\lambda$, we see

$$\mathcal{E}(f,g) = \frac{1}{2} \int \nabla f \nabla g \mathrm{d}\lambda = \frac{1}{2} \int -\Delta f g \mathrm{d}\lambda = \left(-\frac{1}{2}\Delta f,g\right),$$

which directly shows that a half times the Laplace operator Δ is the corresponding generator. We add these correspondences to the ones in figure 1 and we obtain figure 2.



Figure 2: Correspondence between process, semigroup, generator, resolvent and Dirichlet form.

2.3.3 Meaning of the properties

Now that the Dirichlet form is related to an object that we know something about, we can try to see what the definitions that were given before mean and why they make sense. (8) will prove to be a useful tool for this. Let notation be as above: $(\mathcal{E}, \mathcal{D})$ is a Dirichlet form on $L^2(S, \mu)$. The corresponding semigroup is $\{S_t, t \ge 0\}$ with underlying Markov process $X = \{X_t, t \ge 0\}$ and generator (L, D(L)). f, g are functions in $L^2(S, \mu)$.

Meaning of the definition of a symmetric form

First we look at symmetry:

$$\begin{split} \mathcal{E}(f,g) &= \mathcal{E}(g,f) \iff (-Lf,g) = (-Lg,f) \iff (Lf,g) = (Lg,f) \\ \iff (Lf,g) = (f,Lg) \qquad \Longleftrightarrow \ \int Lfg \mathrm{d}\mu = \int fLg \mathrm{d}\mu. \end{split}$$

This shows that the symmetry of \mathcal{E} is directly related to the μ -symmetry of X (see proposition 2.17). We note here that it is possible to extend this theory to non-symmetric forms and processes, but this goes beyond the scope of this report.

Now positivity:

 $\mathcal{E}(f,f) \ge 0 \iff (-Lf,f) \ge 0 \iff (Lf,f) \le 0.$

We see that the positivity of \mathcal{E} is equivalent to the non-positivity of the corresponding generator. This, in turn, comes from the contraction property of the corresponding semigroup on $L^2(S, \mu)$ (prop 2.20). Of course positivity is natural in the context of electrical circuits, since energy dissipation cannot be negative.

Note that (8) clearly implies that \mathcal{E} is linear if and only if L is. In this way we translated all properties of a symmetric form to the corresponding properties of the semigroup or its generator.

The Markov property

Now recall that a Dirichlet form is Markov, which means that if $f \in \mathcal{D}$, then also $g = 1 \land (0 \lor f) \in \mathcal{D}$ and $\mathcal{E}(g,g) \leq \mathcal{E}(f,f)$. It can be shown that this property is equivalent to the sub-Markovness of the semigroup ([5, Theorem 1.4.1]). Since only knowing this fact is not very satisfactory, we can also look at it in a different way. As we mentioned before (in remark 2.26), the Dirichlet form (or Dirichlet energy, as we called it then) can be seen as a measure of how wildly a function varies. The way this is measured depends on the underlying dynamics as the solution to the corresponding Dirichlet problem minimizes this energy. Nothing can be said in general about what this measure of variation should look like, but it must measure variation, so if a function varies less in every way than another function, this measure of variation must be smaller. The Markov property of the Dirichlet form can be shown to be equivalent to the following: whenever $|v(x) - v(y)| \leq |u(x) - u(y)|$ for all x, y and $|v(x)| \leq |u(x)|$ for all x, we have $\mathcal{E}(v) \leq \mathcal{E}(u)$. Without getting into too much detail, this generally says that whenever the distance between the function values of any pair of points is smaller (and v is absolutely smaller than u), then \mathcal{E} must be smaller. This intuitively satisfies the idea that a function that varies less wildly should have a smaller \mathcal{E} value.

Closedness

A Dirichlet form should also be closed, which means that its domain \mathcal{D} should be closed with respect to the metric generated by

$$\mathcal{E}_1(f, f) = \mathcal{E}(f, f) + (f, f) = (-Lf, f) + (f, f).$$

We note here that it can be shown that there exists an operator B such that $B^2 = -L$ (this has all to do with the fact that -L is a positive, symmetric operator). We will write $B =: (-L)^{1/2}$. This square root operator is again symmetric, which means we can write \mathcal{E}_1 as

$$((-L)^{1/2}(-L)^{1/2}f,f) + (f,f) = ((-L)^{1/2}f,(-L)^{1/2}f) + (f,f) = ||(-L)^{1/2}f||^2 + ||f||^2.$$

Using this it is standard to show that \mathcal{D} is complete with respect to the metric induced by \mathcal{E}_1 if and only if $(-L)^{1/2}$ is closed.

Proposition 2.34. $(\mathcal{E}, \mathcal{D})$ is closed if and only if $(-L)^{1/2}$ is closed.

Since L is self-adjoint, so is -L and it can be shown that then also $(-L)^{1/2}$ is self-adjoint. This means that $(-L)^{1/2}$ is closed. With the proposition this means that the closedness of the Dirichlet form is implied by the closedness of the corresponding generator.

The conservative property

As is clear by now, a Dirichlet form is Markov, so it corresponds to a sub-Markov semigroup. However, we can impose a condition such that we get a Markov semigroup. Proposition 2.35. If \mathcal{E} is conservative, then the corresponding semigroup is Markov.

Proof. Let \mathcal{E} be conservative. Then $\mathbb{1} \in \mathcal{D}$ and $\mathcal{E}(\mathbb{1}) = 0$. We already know that the Markov property of the Dirichlet form implies that the semigroup is sub-Markov. It thus suffices to show that $S_t \mathbb{1} = \mathbb{1}$. Let $f \in \mathcal{D}$. Then we have for any $\lambda \in \mathbb{R}$:

$$0 \leq \mathcal{E}(\mathbbm{1} + \lambda f) = \mathcal{E}(\mathbbm{1}) + 2\lambda \mathcal{E}(\mathbbm{1}, f) + \lambda^2 \mathcal{E}(f^2) = 2\mathcal{E}(\mathbbm{1}, f)\lambda + \mathcal{E}(f^2)\lambda^2.$$

This is a quadratic equation in λ and its minimum value is $\frac{-(2\mathcal{E}(\mathbb{1},f))^2}{2\lambda^2}$. Since it cannot be negative, this shows that $\mathcal{E}(\mathbb{1},f)=0$, hence $(-L\mathbb{1},f)=0$. Since this is true for any f in the dense subset \mathcal{D} of $L^2(S,\mu)$, this shows that $L\mathbb{1}=0$ (in $L^2(S,\mu)$). It can be shown that this implies that $S_t\mathbb{1}=\mathbb{1}$ for all t (heuristically: the generator is 0 for $\mathbb{1}$, so S_t is constant for $\mathbb{1}$). \Box

This means that if \mathcal{E} is conservative, there is no loss of mass in the corresponding process (it is conserved) or, in other words, it has infinite lifetime. In the electric circuit context this says: if there is no difference in potential between any pair of nodes, then there is no energy dissipation. This is clearly true, so electrical circuits are conservative.

Irreducibility

The following proposition shows a correspondence between the concepts of irreducibility in different contexts.

Proposition 2.36 . If \mathcal{E} is irreducible, then X is irreducible.

Proof. Suppose X is not irreducible. Then there are disjoint $A_1, A_2 \subset S$ such that $S = A_1 \cup A_2$ and for each *i*: $S_t \mathbb{1}_{A_i} = \mathbb{1}_{A_i}$ for all $t \ge 0$. So in particular

$$L\mathbb{1}_{A_1} = \lim_{t \downarrow 0} \frac{S_t \mathbb{1}_{A_1} - \mathbb{1}_{A_1}}{t} = 0,$$

which implies that $\mathcal{E}(\mathbb{1}_{A_1}) = (-L\mathbb{1}_{A_1}, \mathbb{1}_{A_1}) = 0$. Since for each *i*: $\mu(A_i) > 0$, $\mathbb{1}_{A_1}$ is not identically 0 or 1 in $L^2(S, \mu)$, so there is a non-constant function *f* with $\mathcal{E}(f) = 0$, namely $\mathbb{1}_{A_1}$. This shows that \mathcal{E} is not irreducible.

This concept intuitively makes sense in the context of electrical circuits. The only way in which there can be a non-constant potential without energy dissipation is when the differences in potential are between parts of the circuit that are not connected to each other (meaning that there is infinite resistance along any path from the one part to the other). So the proof above basically says the following. Suppose that X is not irreducible. This means that we can divide the circuit into two parts that are not connected to each other. Then we can give one of the sets potential 1 in each point and the other set potential 0 in each point without any energy dissipation (we use here that the circuit is conservative: on each connected part there is constant potential so no energy dissipation). This means that the energy form is not irreducible.

Regular Dirichlet forms and Hunt processes

In general, the processes corresponding to Dirichlet forms can have unpleasant behaviour, but there is a particular set of 'nice' processes corresponding to regular Dirichlet forms. These processes are called Hunt processes and they are defined as quasi-left-continuous strong Markov processes with cadlag sample paths. We will not elaborate here on what that means. If the Dirichlet form is also local, it corresponds to a diffusion process. We summarize these results in the next theorem.

<u>Theorem 2.37</u>. If \mathcal{E} is regular, X is a Hunt process. If it is also local, X is a diffusion process.

3 Introduction to Riemannian geometry

One of the main themes of this report is random walks and Markov processes on Riemannian manifolds. To be able to study these concepts, we need to introduce some theory about Riemannian manifolds. Appendix A contains a more general introduction to differentiable manifolds. It starts with the definition of a smooth manifold. Then it shows how each point has a tangent space associated to it, which is the space of all vectors tangent to the manifold in that point. Next, it looks at vector fields and discusses differential equations on a manifold. After that it introduces differential forms in order to integrate on a manifold. We will assume all of this to be known here.

This section is about Riemannian manifolds, where we can define the length of a path on the manifold and speak about shortest paths. We will see what 'walking straight ahead' means in a space that is not flat like \mathbb{R}^n , but that is curved. Then we will define connections to be able to move between tangent spaces and we will use this to rigorously define curvature. We end this section with a description of Brownian motion on a manifold and a discussion of the basic operators that are related to Brownian motion. Everything is based on [21] and [12].

3.1 Riemannian manifolds

Let M be a smooth manifold. We assume throughout all of this report that the manifolds that we speak of are *n*-manifolds for some dimension n. We will add some extra structure to it to make it a Riemannian manifold. This will enable us to talk about the norm of tangent vectors and, using that, about lengths of paths. This, in turn, will enable us to define what it means to walk straight forward (locally) on a manifold. However, we first need to treat some theory involving vector spaces and inner products. In all of this we follow [21].

3.1.1 Inner product spaces

Inner product

Let us begin with the definition of an inner product. Throughout this report we will assume that the vector spaces we work with have finite dimension, usually denoted by n.

<u>Definition 3.1</u>. An *inner product* on a vector space V is a non-zero bilinear function $\langle , \rangle : V \times V \to \mathbb{R}$ such that

- $\langle v, w \rangle = \langle w, v \rangle$ $\forall v, w \in V$ (symmetric)
- $\langle v, v \rangle > 0$ $\forall v \neq 0$ (positive-definite).

Now let $v_1, ..., v_n$ be a basis for V and let $v_1^*, ..., v_n^*$ be the corresponding dual basis of V^* . Then it is easy to see that we can write

$$\langle , \rangle = \sum_{i,j=1}^n g_{ij} v_i^* \otimes v_j^*,$$

where

$$g_{ij} = \langle v_i, v_j \rangle.$$

Denote by (g_{ij}) the matrix with elements g_{ij} for rows i = 1, ..., n and columns j = 1, ..., n. Then the symmetry of \langle , \rangle implies that (g_{ij}) is symmetric.

Given an inner product we can define a map $\alpha : V \to V^*$ that takes v to $(w \mapsto \langle v, w \rangle)$. α can be shown to be an isomorphism. If we write α as a matrix with respect to the bases $v_1, ..., v_n$ and $v_1^*, ..., v_n^*$ as defined above, this yields (g_{ij}) again. This implies that $\det(g_{ij}) \neq 0$.

The positive definiteness of the inner product implies that (g_{ij}) is also positive definite. We can use α to define an inner product on V^* . We define for $\lambda, \mu \in V^*$:

$$\langle \lambda, \mu \rangle^* := \mu(\alpha^{-1}(\lambda).$$

It can be checked that \langle , \rangle^* is indeed an inner product on V^* . It turns out that we can write is as

$$\langle , \rangle^* = \sum_{i,j=1}^n g^{ij} v_i \otimes v_j,$$

(through the identification $V = V^{**}$) where the g^{ij} are such that $(g^{ij}) = (g_{ij})^{-1}$.

<u>Definition 3.2</u>. We call a base $v_1, ..., v_n$ for V orthonormal if $\langle v_i, v_j \rangle = \delta_i^i$.

The inner product \langle , \rangle^* can be characterized as the inner product on V^* such that $v_1, ..., v_n$ is orthonormal with respect to \langle , \rangle if and only if its dual basis $v_1^*, ..., v_n^*$ is orthonormal with respect to \langle , \rangle^* .

Norm

We can use the inner product to define a norm. We define for $v \in V : ||v|| = \langle v, v \rangle^{1/2}$. This can be shown to satisfies the norm properties:

- $||av|| = |a| \cdot ||v||$ for $a \in \mathbb{R}$ and $v \in V$
- $||v + w|| \le ||v|| + ||w||$ (triangle inequality)
- $||v|| \ge 0$ and ||v|| = 0 if and only if v = 0

Moreover, we have the Cauchy-Schwarz Inequality: $|\langle v, w \rangle| \leq ||v|| \cdot ||w||$. There is equality if and only if v is a multiple of w or vice versa.

Unitary alternating functions

Finally we briefly study the vector space $\Omega^k(V)$ of alternating, linear functions on V^k . We will take k = n, the dimension of V. Then $\Omega^n(V)$ has dimension 1. It can be shown that if $v_1, ..., v_n$ and $w_1, ..., w_n$ are orthonormal bases for V, then

$$v_1^* \wedge \ldots \wedge v_n^* = \pm w_1^* \wedge \ldots \wedge w_n^*.$$

We call these the unitary elements of $\Omega^n(V)$. If $u_1, ..., u_2$ is any other base for V and we write $u_i = \alpha^{ji} v_j$ and $A = (\alpha^{ij})$, then it can be shown that

$$\det(A)u_1^* \wedge \ldots \wedge u_n^* = v_1^* \wedge \ldots \wedge v_n^*$$

Write

$$\langle z_1, z_2 \rangle = \sum_{i,j=1}^n g_{ij} u_i^* \wedge u_j^*(z_1, z_2) \text{ for } z_1, z_2 \in V,$$

where $g_{ij} = \langle u_i, u_j \rangle$. Also denote $G = (g_{ij})$. Then $G = A^T A$, so $\det(G) = \det(A)^2 (> 0)$, which shows that

$$\sqrt{\det(g_{ij})u_1^* \wedge \ldots \wedge u_n^*} = v_1^* \wedge \ldots \wedge v_n^*.$$

Since the unitary elements are $\pm v_1^* \wedge .. \wedge v_n^*$, we see that the unitary elements can be written with respect to $u_1, .., u_n$ as

$$\pm \sqrt{\det(g_{ij})} u_1^* \wedge \ldots \wedge u_n^*.$$

If M is orientable and has orientation μ , we define the unitary element that returns a positive number when applied to any $v_1, ..., v_n \in \mu$ to be the positive unitary element.

3.1.2 Riemannian metric

Riemannian metric

We can now define the important, additional structure on a manifold that we will use here.

<u>Definition 3.3</u>. Let M be a smooth manifold and TM its tangent bundle. We call a map \langle , \rangle a C^{∞} Riemannian metric for TM if

- for any $p \in M$, \langle , \rangle_p is an inner product on $T_p M$
- for any two sections s_1, s_2 , the function $p \mapsto \langle s_1(p), s_2(p) \rangle$ is C^{∞} .

We can now locally apply the theory of vector spaces described above. Let (x, U) be a coordinate system for M. Then on U we can write

$$\langle , \rangle = \sum_{i,j=1}^{n} g_{ij} \mathrm{d} x^{i} \otimes \mathrm{d} x^{j}.$$

Then (g_{ij}) is symmetric, C^{∞} and has $\det(g_{ij}) > 0$. \langle , \rangle induces a Riemannian metric \langle , \rangle^* on the dual bundle T^*M and we can write it as

$$\langle \,,\rangle^* = \sum_{i,j=1}^n g^{ij} \partial_i \otimes \partial_j,$$

where we identify $(T_p M)^{**}$ with $T_p M$ in the usual way. As above, for each p these g^{ij} satisfy $(g^{ij})(p) = (g_{ij})^{-1}(p)$.

Note that $g_{ij} = \langle \partial_i, \partial_j \rangle$. Since ∂_i and ∂_j are smooth sections of TM on U, definition 3.3 tells us that $p \mapsto g_{ij}(p) = \langle \partial_i(p), \partial_j(p) \rangle_p$ is a smooth function on U. Using the same argument with T^*M yields that q^{ij} is also smooth on U.

Change of coordinates

Suppose (x, U) and (y, V) are coordinate systems around some point $p \in M$. Denote the matrix of the metric in x-coordinates G and in y-coordinates \hat{G} . Then we see

$$\hat{g}_{ij} = \left\langle \frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j} \right\rangle = \left\langle \frac{\partial x^k}{\partial y^i} \frac{\partial}{\partial x^k}, \frac{\partial x^l}{\partial y^j} \frac{\partial}{\partial x^l} \right\rangle = \frac{\partial x^k}{\partial y^i} \frac{\partial x^l}{\partial y^j} \left\langle \frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l} \right\rangle = \frac{\partial x^k}{\partial y^i} \frac{\partial x^l}{\partial y^j} g_{kl}$$

If we denote the Jacobian matrix $\left(\frac{\partial x^i}{\partial y^j}\right)_{i,j=1}^n$ by J, we obtain $\hat{G} = J^T G J$. This also shows that $\hat{G}^{-1} = J^{-1} G^{-1} (J^T)^{-1} = J^{-1} G^{-1} (J^{-1})^T$.

Raising and lowering indexes

The map α that we described above is a special case of what we call lowering indexes. We can transform any vector to a covector by mapping v to v^{\flat} (v flat) defined by $w \mapsto \langle v, w \rangle$. In local coordinates (x, U), this means that

$$v^{\flat}(w) = \left\langle v^{i}\partial_{i}, w^{j}\partial_{j} \right\rangle = v^{i}w^{j}g_{ij} = g_{ij}v^{i}\mathrm{d}x^{j}(w), \text{ so } v^{\flat} = g_{ij}v^{i}\mathrm{d}x^{j}.$$

If we write $v^{\flat} = v_j dx^j$ (as usual with covectors), we see that we have lowered the j (hence the \flat notation) and that $v_j = g_{ij}v^i$. Since g is invertible, we can also define its inverse map (sharp) that takes a covector $\omega = \omega_i dx^i$ to the vector $\omega^{\sharp} = g^{ij}\omega_i\partial_j$, so the new components are $\omega^j = g^{ij}\omega_i$ and we say that the index has been raised.

It is possible to generalize these operators to apply them to tensors, although we have to specify precisely which index should be lowered or raised. For instance, if $F: V \times V^* \times V \to \mathbb{R}$ has coefficients $F_i^{j}{}_k$, then we can raise the first coordinate and obtain F^{\sharp} given by $F^{ij}{}_k = g^{il}F_l^{j}{}_k$.

Norm and curve length

Since we have an inner product on the tangent spaces T_pM , this induces a norm on T_pM . We define for $v \in T_pM$:

$$||v|| = \langle v, v \rangle_n^{1/2}$$

This gives us a way to define lengths of curves in M.

<u>Definition 3.4</u>. Let $\gamma : [a, b] \to M$ be C^{∞} . We define the *length* of γ from a to b as

$$L_a^b(\gamma) = \int_a^b ||\gamma'(t)|| \mathrm{d}t.$$

If γ is only piecewise C^{∞} (but still continuous), we define its length as the sum of the lengths of the smooth parts.

This definition agrees with the usual length of a curve in \mathbb{R}^n . Usually a and b are left out of the notation. We can define the arc-length function $s(t) = L_a^t(\gamma)$. For any curve we can shift its parametrization such that a = 0. Then we say that γ is parametrized by arc length if s(t) = t for all $t \in [0, b - a]$.

Induced metric

Now suppose M is connected. This means that any two points are connected by a smooth curve. We define for any $p, q \in M$:

 $d(p,q) = \inf\{L(\gamma) : \gamma \text{ is a piecewise smooth curve from } p \text{ to } q\}.$

It can easily be shown that this defines a metric on M. In fact, (M, d) is homeomorphic to M equipped with its original metric (the one with respect to which it is a manifold), so the metrics induce the same topology.

3.1.3 Integration

Volume element

If (x, U) is some chart and $p \in U$, we can write the unitary elements of $\Omega^n(T_pM)$ as

$$\pm \sqrt{\det(g_{ij})(p)} \mathrm{d}x^1 \wedge .. \wedge \mathrm{d}x^n(p).$$

If M is orientable we can define one of them as the positive unitary element and if (x, U) is orientation preserving, we can write this element on U as

$$\mathrm{d}V = \sqrt{\mathrm{det}(g_{ij})} \mathrm{d}x^1 \wedge \ldots \wedge \mathrm{d}x^n.$$

<u>Definition 3.5</u>. We call dV the volume element determined by \langle , \rangle and we call

$$\int_M \mathrm{d}V$$

the volume of M.

Note that the volume of M might be infinite. It is also possible to construct a volume element on a non-orientable manifold, but we will not do that here.

Measure theoretic considerations

In the appendix, we define integration of a k-form over a manifold by pulling it back to a subset of \mathbb{R}^n and using the usual interpretation of integration there. However, this is a rather indirect definition and in a lot of cases it would be nice to view the volume as a measure on the manifold.
Then we can use all results that we know from measure theory on this manifold. This turns out to be possible on a Riemannian manifold, as is described in [7, Section 3.4].

Define the Borel σ -algebra $\mathcal{B}(M)$ as the smallest σ -algebra containing all open subsets of M. Further call a set V measurable if for any coordinate system $(x, U) \ x(U \cap V)$ is a Lebesgue measurable subset of \mathbb{R}^n . The measurable sets form a σ -algebra, which we denote by $\Lambda(M)$. If V is any open set and (x, U) is a coordinate system, then $U \cap V$ is also open, so $x(U \cap V)$ is open and hence Lebesgue-measurable. This implies that $\mathcal{B}(M) \subset \Lambda(M)$. The following theorem guarantees us the existence of a canonical measure on M that is defined on $\Lambda(M)$.

<u>Theorem 3.6 ([7, Thm 3.11])</u>. Let M be any Riemannian manifold. Then there exists a measure ν on $\Lambda(M)$ such that in any chart (x, U): $d\nu = \sqrt{\det G} d\lambda$. Here G and λ are the metric matrix and Lebesgue measure in (x, U), respectively.

We see from the formula in local coordinates that integration with respect to ν in the theorem is the same as the integration of the volume element (apart from the fact that for V the orientation matters, note that we did not assume orientability here). We will therefore simply write $\nu = V$ and treat V as a measure on M. The following proposition gives us some properties of V.

Proposition 3.7 ([7, Thm 3.11]). V satisfies the following properties.

- V is complete.
- $V(K) < \infty$ for any compact $K \subset M$.
- V(O) > 0 for any open, non-empty $O \subset M$.
- (inner regularity) For any $A \in \Lambda(M)$: $V(A) = \sup\{V(K) : K \subset A, K \text{ compact}\}.$
- (outer regularity) For any $A \in \Lambda(M)$: $V(A) = \inf\{V(O) : A \subset O, O \text{ open}\}$.

3.1.4 Geodesics

Shortest curves between two points

We have seen that we can define the distance between two points as the infimum of the lengths of all piecewise smooth curves from one point to the other. However, there is not always a piecewise smooth curve from p to q with length d(p,q). As an easy example take $M = \mathbb{R}^2 \setminus \{0\}$ and p = -q. Every piecewise smooth curve between them has to deviate a bit from the straight line from -p to p close to 0, but this deviation can be arbitrarily small. So $d(p, -p) = 2||p||_2$, where $||\cdot||_2$ denotes the Euclidean norm, but every piecewise smooth curve has a larger length. It is also possible to have many shortest paths. For example from the north pole to the south pole of a (perfectly round) globe, there are infinitely many shortest paths (all meridians). We would like to find shortest curves between points and examine when they exist. To find a curve that minimizes the length functional L, we have to use the calculus of variations. We will briefly explain the idea behind this and introduce some related concepts. Then we will get to the results.

Variations

When trying to find the minimum of a smooth function $f : \mathbb{R} \to \mathbb{R}$, one usually calculates the derivative and checks for which points it is 0. These points are not necessarily the points where a minimum is attained, but they are the only candidates. This same kind of strategy can be used when minimizing over curves. Let $\gamma : [a, b] \to M$ be a smooth curve with $\gamma(a) = p$ and $\gamma(b) = q$.

<u>Definition 3.8</u>. A variation of γ is a function $\alpha : (-\epsilon, \epsilon) \to M$ for some $\epsilon > 0$ such that:

- $\alpha(0,t) = \gamma(t)$
- there are $a = t_0 < t_1 < ... < t_N = b$ such that α is C^{∞} on every set $(-\epsilon, \epsilon) \times [t_{i-1}, t_i]$.

If, additionally, $\alpha(u, a) = p$ and $\alpha(u, b) = q$ for all $u \in (-\epsilon, \epsilon)$, we way that α is a variation of γ keeping endpoints fixed.

The idea is that for each $u \in (-\epsilon, \epsilon)$, the curve $t \to \alpha(u, t)$ is a curve that is very close to α . We will denote these curves by $\bar{\alpha}(u)$. We will consider a curve γ a candidate for a minimal curve whenever

$$\frac{\mathrm{d}L(\bar{\alpha}(u))}{\mathrm{d}u}|_{u=0} = 0,$$

for all variations α of γ that keep endpoints fixed. Instead of minimizing the length functional, we can minimize the energy functional $E(\gamma) = \frac{1}{2} \int_a^b ||\gamma'(t)||^2 dt$. This can be shown to be equivalent as long as we consider curves that are parametrized by a constant plus another constant times arc length (so curves with constant speed). Since the parametrization of a curve does not influence its length, this is not a real restriction.

Differential equations for geodesics

Now we skip a lot of calculations and jump right to the result.

 $\frac{\text{Proposition 3.9}}{L = L_a^b \text{ if and only if } \gamma \text{ is } C^\infty \text{ on } [a,b] \to M \text{ such that } \frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = 0 \text{ is a critical point for } k = 1,..,n:$

$$\frac{\mathrm{d}^2 \gamma^k}{\mathrm{d}t^2} + \sum_{i,j=1}^n \Gamma^k_{ij}(\gamma(t)) \frac{\mathrm{d}\gamma^i}{\mathrm{d}t} \frac{\mathrm{d}\gamma^j}{\mathrm{d}t} = 0, \qquad (9)$$

when $\gamma(t) \in U$. Here

$$\Gamma_{ij}^{k} = \sum_{l=1}^{n} g^{kl}[ij,l] = \sum_{l=1}^{n} g^{kl} \frac{1}{2} \left(\frac{\partial g_{il}}{\partial x^{j}} + \frac{\partial g_{jl}}{\partial x^{i}} - \frac{\partial g_{ij}}{\partial x^{l}} \right)$$

and (g^{ij}) and (g_{ij}) are the coefficients of \langle , \rangle resp \langle , \rangle^* with respect to the coordinate system (x, U).

We will not give a proof here, but it is good to mention that the equations in the proposition are exactly the Euler-Lagrange equations for the energy functional.

<u>Definition 3.10</u>. We call the curves from the previous proposition *geodesics* and we call (9) the *geodesic equations*.

Note that any reparametrization of a geodesic also minimizes the length functional (but we do not call them geodesics).

Local existence

Now we can ask ourselves the question if such geodesics exist. Let us first look at the easiest example: \mathbb{R}^n . Here the matrix (g_{ij}) is the identity (and hence constant), so all derivatives of g_{ij} are 0, so Γ_{ij}^k is 0 for any i, j, k. This means that the geodesic equations are reduced to $\frac{d^2\gamma^k}{dt^2} = 0$, so every component of the geodesic should be of the form $t \mapsto a + bt$. This means that the geodesic is a line in \mathbb{R}^n .

Now we give a result for the general case. It is shown in section A.3 how to solve a simple differential equation on a manifold. The geodesic equations are obviously more complex, but we can still solve this equation. Since it is a second order differential equation, we expect that we need an additional initial condition. This turns out to be correct, as we see in the next proposition.

Proposition 3.11. Let p in M. Then there is neighbourhood U of p and an $\epsilon > 0$ such that for each point $q \in U$ and each tangent vector $v \in T_q M$ such that $||v|| < \epsilon$, there exists a unique geodesic $\gamma_v : (-2, 2) \to M$ such that $\gamma_v(0) = q$ and $\gamma'_v(0) = v$. In particular, this means that there is always a neighbour of a point p such that there is a unique geodesic through p in that neighbourhood with a specified tangent vector at p.

Global existence

We have now seen some local behaviour (some more will follow when studying the exponential map). To speak about global properties we introduce the following concept.

<u>Definition 3.12</u>. We call a Riemannian manifold M geodesically complete if any geodesic γ : $[a,b] \to M$ can be extended to a geodesic $\mathbb{R} \to M$.

This is what we would like. That geodesics do not only have nice properties locally, but that any geodesic can be followed for as long as we wish. The next theorem tells us exactly when this is the case.

Theorem 3.13 (Hopf-Rinow-De Rahm). Let M be a Riemannian manifold with Riemannian metric \langle , \rangle . Then M is geodesically complete if and only if it is complete with respect to the metric determined by \langle , \rangle . Moreover, when this is the case, any two points can be joined by a geodesic of minimal length.

If M is compact with respect to its original metric, then it is also compact with respect to the metric d induced by \langle , \rangle (recall that M with its original metric is homeomorphic to (M, d)). This means it is complete with respect to d, so the results of the theorem hold for any compact manifold.

3.1.5 Exponential map

Following geodesics

Now that we know that geodesics exist (at least locally), we can follow them. This provides a natural map from the tangent space at a point to the manifold itself.

<u>Definition 3.14</u>. Let $p \in M$ and $v \in T_pM$ and suppose that there is a geodesic $\gamma : [0,1] \to M$ such that $\gamma(0) = q$ and $\gamma'(0) = v$. Then we define $\exp(v) := \exp_p(v) := \gamma(1)$ and we call it the *exponential* of v.

Note that $\exp_p(v)$ is just the point where one ends up after following γ for one unit time. We can now describe γ as $\gamma(t) = \exp_p(tv)$. Note that we have $\frac{d}{dt}|_0 \exp_p(tv) = \frac{d}{dt}|_0\gamma(t) = v$, which we are used to when dealing with exponential maps (the full reason that the map is called exponential has to do with Lie algebras, so we will not discuss it here). In \mathbb{R}^n with the Euclidean metric the exponential map takes a special form. Suppose $p \in \mathbb{R}^n$ and $v \in T_p \mathbb{R}^n = \mathbb{R}^n$. Geodesics in \mathbb{R}^n are just straight lines. This means that after following the geodesic with tangent vector v at pfor one time unit, one ends up in p + v. Hence $\exp_p(tv) = p + tv$ for any $t \in \mathbb{R}$. This is a very special case, since the tangent space can be identified with the space itself (as a vector space).

Some properties

We will study the exponential map more closely, since we will be working with this map later. The results are basically [21, Thm 9.14-9.18]. The following properties further characterize the local behaviour of geodesics and the exponential map.

Proposition 3.15. Let $p \in M$. Then there is a neighbourhood W of p and an $\epsilon > 0$ such that the following statements hold.

- Any two points in W are joined by a unique geodesic of length $< \epsilon$.
- If v(q,q') denotes the (unique) vector in $v \in T_qM$ of length $< \epsilon$ such that $\exp_q(v) = q'$, then $(q,q') \mapsto v(q,q')$ is C^{∞} as a map $W \times W \to TM$.

- For each $q \in W$, \exp_q maps the open ball of radius ϵ around the origin of $T_q M$ diffeomorphically onto an open set U_q that contains W.
- For every $0 < c < \epsilon$, the geodesics through q are perpendicular to the hypersurfaces $\{\exp_q(v) : ||v|| = c\}.$
- If $\gamma : [0,1] \to M$ is the geodesic of length $\langle \epsilon \rangle$ between q and q' and $c : [0,1] \to M$ is any piecewise smooth curve from q to q', then $L(\gamma) \leq L(c)$. Equality only holds if c is a reparametrization of γ .

In particular we see that any point has a neighbourhood on which the (uniquely defined) geodesics are in fact the shortest paths (up to reparametrization).

3.1.6 Normal coordinates

Normal neighbourhood

It will often prove useful to consider a specific coordinate system around a point. To be able to define these coordinates (following [12]), we first need the following.

<u>Definition 3.16</u>. Let V be a vector space and $v \in V$. We call a neighbourhood U of v star-shaped if for any $w \in U$ and any $\lambda \in [0, 1]$: $\lambda v + (1 - \lambda)w \in U$.

This means that for any point of U it should contain the line from that point to v.

<u>Definition 3.17</u>. We call a neighbourhood U of p normal if it is the diffeomorphic image of a star-shaped neighbourhood of $0 \in T_p M$ under \exp_p .

The following lemma guarantees us that such neighbourhoods exist (by considering star-shaped subsets of V).

<u>Lemma 3.18</u>. Every $p \in M$ has a neighbourhood U that is the diffeomorphic image of a neighbourhood V of $0 \in T_p M$ under \exp_p .

Normal coordinates

Observe that if $e_1, ..., e_n$ is an orthonormal basis for T_pM , we can make a natural mapping E of a vector $v \in T_pM$ to its coordinates with respect to the basis. It is easy to show that this mapping is a vector space isomorphism (basically because $\langle \partial_i, \partial_j \rangle = \delta_i^i$).

<u>Definition 3.19</u>. Let U be a normal neighbourhood of p and let $e_1, ..., e_n$ be an orthonormal basis for T_pM with mapping $E: T_pM \to \mathbb{R}^n$ as above. Define

$$x: M \to \mathbb{R}^n$$
 by $x = E \circ \exp_p^{-1}$.

Then (x, U) is a coordinate system. We call this system normal coordinates centered at p.

Note that x(p) = 0, so it is natural to call x a coordinate system centered at p. A lot of things involving the metric and geodesics simplify a lot with respect to this coordinate system.

Proposition 3.20 ([12, Prop 5.11]). Let (x, U) be a normal coordinate system centered at $p \in M$. Then the following assertions hold.

- The geodesic from p in the direction of $v = v^i \partial_i$ is given by the line $t \mapsto (tv^1, ..., tv^n)$ with respect to x (as long as it is in U).
- $g_{ij} = \delta^i_j$ (so G = I) at p
- Any Euclidean ball $\{p \in M : \sum x^i(p)^2 < c\}$ contained in U is the diffeomorphic image of a ball around $0 \in T_p M$ under \exp_p .

• The first partial derivatives of g_{ij} and the Christoffel symbols are 0 at p.

Note that geodesics through U that do not visit p generally are not a straight line with respect to x. Also, the metric matrix and the Christoffel symbols only vanish at p itself, but generally not on a neighbourhood of p. The latter is only the case when that neighbourhood is isometric to an open subset of \mathbb{R}^n , which is a very strong assumption (see the end of section 3.2).

3.2 Connections and curvature

As we have seen a lot of times now, it is convenient to work with a smooth manifold by using its tangent vectors. In each point, the tangent vectors form a vector space, so we have many tools to work with them. However, if we would like to compare vectors that are tangent to different points, we have an immediate difficulty: they are elements of different vector spaces. We can also not simply identify the vector spaces with each other. In order to deal with this, we will introduce connections and parallel transport. Using these important notions, we can then rigorously define the curvature of a manifold. In all of this section we will follow [12].

3.2.1 Connections

We will discuss connections, derivatives along curves and parallel transport. For each of these, we first define it in the general setting of a smooth manifold and then we show the properties corresponding to the natural version in a Riemannian manifold. The main idea is that we would like to define the derivative of a vector field along a curve. Then we can say that the vectors are in some way 'constant' along that curve if the derivative is 0. This will provide us with a way to compare vectors from different tangent spaces.

Differentiating vector fields

Let us start with the central definition.

<u>Definition 3.21</u>. Let M be a smooth manifold and let $\pi : E \to M$ be a vector bundle over M. Denote the set of smooth sections of E by $\mathcal{E}(M)$. Then a connection in E is a map $\nabla : \mathscr{T}(M) \times \mathcal{E}(M) \to \mathcal{E}(M)$ that maps (X, Y) to $\nabla_X Y$ in such a way that

- $\nabla_X Y$ is linear over C^{∞} in X
- $\nabla_X Y$ is linear over \mathbb{R} in Y
- (product rule) $\nabla_X(fY) = f\nabla_X Y + (Xf)Y$ for $f \in C^\infty$.

We also call $\nabla_X Y$ the covariant derivative of Y in the direction of X. In particular, if we use the tangent bundle the connection reduces to a map $\mathscr{T}(M) \times \mathscr{T}(M) \to \mathscr{T}(M)$ and we call it a *linear connection*.

The intuition is that in every point p we take the derivative of Y in the direction of the vector X_p . This suggests that we only need to know X_p and the values Y_q for q in some neighbourhood of p to calculate $\nabla_X Y$. This turns out to be true.

Lemma 3.22 ([12, Lemma 4.2]). Suppose ∇ is a connection in a bundle E over M. Let X be a vector field and let Y be a section of E. Then for any point $p \in M$, the value $\nabla_X Y|_p$ only depends on X_p and the values of Y in a neighbourhood of p.

Explanation of the definition

We will now explain the defining properties of a connection more carefully. First of all, since $\nabla_X Y|_p$ is a derivative in the direction of X_p , it should depend linearly (over \mathbb{R}) on X_p . Since this can be done in every point and these points do not influence each other, $\nabla_X Y$ should

depend linearly (over C^{∞}) on X.

Now suppose that Y is multiplied by a C^{∞} function f. The reasoning above does not work, since $\nabla_X Y|_p$ does not only depend on Y_p . Locally in p, Y is multiplied by a constant f(p), so this should appear in front of $\nabla_X Y$. However, it also matters how f varies around p, which gives an extra term. Together this forms the product rule from the definition. In particular, if f is constant, the extra term vanishes and the expression is reduced to a constant times $\nabla_X Y$, which is a special case of the second requirement. Also, naturally the derivative of the sum should be the sum of the derivatives. This gives us the full second requirement.

In the following we will assume that we work with the tangent bundle, so a connection is always a linear connection.

Generalized Christoffel symbols

Suppose (x, U) is some coordinate chart around p and we regard the vector fields $X = \partial_i$ and $Y = \partial_j$ on U (so $X_p = \frac{\partial}{\partial x^i}|_p$). Let ∇ be a connection. Since $\nabla_X Y|_p$ only depends locally on X and Y, we can calculate its value in p. We can use this to define the coefficients C_{ij}^k by

$$\nabla_{\partial_i}\partial_j = C_{ij}^k \partial_k. \tag{10}$$

The C_{ij}^k can be considered generalized versions of the Christoffel symbols. We will come back to this shortly.

A connection is fully described by the C_{ij}^k . Using just the defining properties, it is easy to show that

$$\nabla_X Y = (XY^k + X^i Y^j C^k_{ij})\partial_k.$$

It also works the other way around: any choice of smooth functions C_{ij}^k locally defines a connection. By a partition of unity argument, this can be used to define a connection on any manifold.

Levi-Civita connection

We have seen that a variety of connections can be constructed. On a Riemannian manifold there is a distinguished one, called the Levi-Civita connection. It satisfies some extra properties, which we will introduce first.

<u>Definition 3.23</u>. Let ∇ be a (linear) connection on M. We have the following definitions.

- ∇ is called *compatible with the metric* if for all vector fields $X, Y, Z : X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle$.
- ∇ is called *symmetric* if $\nabla_X Y \nabla_Y X = [X, Y] (:= XY YX).$

The following important theorem tells us that a connection with such properties uniquely exists.

Theorem 3.24 ([12, Thm 5.4], Fundamental Lemma of Riemannian Geometry). Let M be a Riemannian manifold. Then there is a unique connection that is symmetric and compatible with the metric. We call this the Levi-Civita connection.

In the case of the Levi-Civita connection, the C_{ij}^k are just the Christoffel symbols. In fact, once uniqueness is shown in the proof of the theorem (using a lot of juggling with equations), one simply sets $C_{ij}^k = \Gamma_{ij}^k$ and checks that the corresponding connection is symmetric and compatible with the metric. In particular, in \mathbb{R}^n the Christoffel symbols are 0, so the Levi-Civita connection is $\nabla_X Y = (XY^k)\partial_k$. This is also called the Euclidean connection.

Curves

Now that we can take derivatives of vector fields with respect to vector fields, we would like to define how to do this with respect to curves. We therefore need the following definitions.

<u>Definition 3.25</u>. A curve in M is a smooth map $\gamma : I \to M$ for some interval I. A vector field along a curve $\gamma : I \to M$ is a smooth map $V : I \to TM$ such that $V(t) \in T_{\gamma(t)}M$. It is called extendible if there is a vector field \tilde{V} defined on a neighbourhood of the image of γ such that $V(t) = \tilde{V}(\gamma(t))$. We denote the set of all vector fields along γ by $\mathscr{T}(\gamma)$.

For any smooth curve γ one can define its velocity vector $\dot{\gamma} = \frac{d}{dt}\gamma$ in each point. This is an example of a vector field along a curve. The notion of extendibility is necessary to link differentiation along a curve to differentiation with respect to vector fields, as becomes clear in the following lemma.

Derivative along a curve

Lemma 3.26 ([12, Lemma 4.9]). Let ∇ be a connection on M and let $\gamma : I \to M$ be a curve. Then there is a unique operator $D_t : \mathscr{T}(\gamma) \to \mathscr{T}(\gamma)$ such that:

- D_t is linear over \mathbb{R} .
- (product rule) $D_t(fV) = \dot{f}V + fD_tV$ for $f \in C^{\infty}(I), V \in \mathscr{T}(\gamma)$
- For any $V \in \mathscr{T}(\gamma)$ and any extension \tilde{V} of V: $D_t V(t) = \nabla_{\dot{\gamma}(t)} \tilde{V}$.

We call $D_t V$ the covariant derivative of V along γ .

Note that we want linearity as a basic property of differentiation. Further, the product rule looks a lot like the one for connections. The last requirement links the D_t to ∇ in a natural way. The proof first shows that the value $D_t V$ at t_0 only depends on V in a neighbourhood of t_0 . This allows us to use local coordinates to calculate $D_t V$. Let (x, U) be coordinates around $\gamma(t_0)$. Then we can write $V(t) = V^j(t)\partial_j$. The ∂_j are extendible, so we can use the defining properties of $D_t V$ to see:

$$D_{t}V(t_{0}) = D_{t}(V^{j}\partial_{j})(t_{0}) = \dot{V}^{j}(t_{0})\partial_{j} + V^{j}(t_{0})\nabla_{\dot{\gamma}(t_{0})}\partial_{j}$$

$$= \dot{V}^{j}(t_{0})\partial_{j} + V^{j}(t_{0})\nabla_{\dot{\gamma}^{i}(t_{0})\partial_{i}}\partial_{j} = \dot{V}^{j}(t_{0})\partial_{j} + V^{j}(t_{0})\dot{\gamma}^{i}(t_{0})\nabla_{\partial_{i}}\partial_{j}$$

$$= \dot{V}^{j}(t_{0})\partial_{j} + V^{j}(t_{0})\dot{\gamma}^{i}(t_{0})C_{ij}^{k}(\gamma(t_{0}))\partial_{k} = (\dot{V}^{k}(t_{0}) + V^{j}(t_{0})\dot{\gamma}^{i}(t_{0})C_{ij}^{k}(\gamma(t_{0})))\partial_{k}$$
(11)

This equation shows uniqueness (since the expression does not depend on D anymore) and it can be used as a definition to prove existence, but it will also be important later.

Geodesics

We will use derivatives along a curve to give another approach to geodesics. To do this we first need the following definition.

<u>Definition 3.27</u>. Let γ be a curve. We define its *acceleration* (with respect to ∇) as the derivative of its velocity field along itself: $D_t \dot{\gamma}$.

We now have an alternative way of speaking about straight lines in a curved space. We can call a curve straight precisely when its acceleration is 0, then the velocity vector 'stays the same' along the curve (we will soon make this precise).

<u>Definition 3.28</u>. We call a curve a *geodesic* (with respect to ∇) if its acceleration is 0.

Note that this is a more general definition of a geodesic: we have not used the Riemannian structure. However, we must show that it coincides with our previous definition.

Riemmanian geodesics

Suppose again that we have a Riemannian metric and let ∇ be the Levi-Civita connection. Then $C_{ij}^k = \Gamma_{ij}^k$ (for any coordinate chart (x, U) around p). Let γ be a smooth curve. Then γ is a geodesic if $D_t \dot{\gamma} \equiv 0$. Using (11), this means that for all k and all t_0 (and in any coordinate system):

$$\ddot{\gamma}^{k}(t_{0}) + \dot{\gamma}^{j}(t_{0})\dot{\gamma}^{i}(t_{0})\Gamma^{k}_{ij}(\gamma(t_{0})) = 0.$$

In other words, γ is a geodesic (for the Levi-Civita connection) if and only if it satisfies the geodesic equations.

Parallel transport

We have mentioned a few times that we consider a vector field 'constant' along a curve if the derivative along the curve is 0. We will formalize this in the following definition.

<u>Definition 3.29</u>. Let V be a vector field along some curve γ . We say that V is *parallel* along γ (with respect to ∇) if $D_t V = 0$. If a vector field V on M is parallel along any curve, we simply call it *parallel*.

The following theorem shows an important property.

Theorem 3.30 ([12, Thm 4.11], Parallel translation). Let $\gamma : I \to M$ be a curve and let $t_0 \in I$. For any vector $V_0 \in T_{\gamma(t_0)}M$ there exists a unique parallel vector field V along γ such that $V(t_0) = V_0$.

The idea is that V_0 is moved along γ while being kept parallel with respect to itself in the previous positions (which is why we speak of parallel translation). The vector field that is obtained is called the parallel translate of V_0 along γ . We can now define an operator $P_{t_0t_1}: T_{\gamma(t_0)}M \to T_{\gamma(t_1)}M$ that maps a vector V_0 at $\gamma(t_0)$ to $V(t_1)$, where V is the parallel translate of V_0 along γ .

Comparing vectors

We have found a way to transport vectors, while keeping them parallel to the initial vector. To compare vectors from different tangent spaces, we can simply transport one of them over a suitable curve to the other vector space and compare the vectors there. In this way we can obtain another way of interpreting differentiation along a curve.

Proposition 3.31. Let ∇ be a connection on M and let V be a vector field along a curve γ . Then the following holds:

$$D_t V(t_0) = \lim_{t \to t_0} \frac{P_{t_0 t}^{-1} - V(t_0)}{t - t_0}$$

Now D_t really takes the form of a derivative. The subtraction is well defined since $P_{t_0t}^{-1}$ is in the same vector space as $V(t_0)$.

The Riemannian situation

When M has a Riemannian metric, there is even extra structure.

<u>Proposition 3.32</u>. Let M be a Riemannian manifold and ∇ the corresponding Levi-Civita connection. Then the following statements hold.

- If V, W are vector fields along some curve γ , then $\frac{d}{dt} \langle V, W \rangle = \langle D_t V, W \rangle + \langle V, D_t W \rangle$.
- If V, W are both parallel along γ , then $\langle V, W \rangle$ is constant along γ .
- The parallel translation operator $P_{t_0t_1}$ is an isometry $T_{\gamma(t_0)}M \to T_{\gamma(t_1)}M$.

The properties above are actually equivalent to the fact that ∇ is compatible with the metric. This is all [12, Lemma 5.2].

3.2.2 Curvature

Intrinsic and extrinsic curvature

One of the most important differences between a general manifold and \mathbb{R}^n is that a manifold does not have to be flat: it can be curved. However, it is not directly clear how to quantify this and it is even not so straightforward what curvature means. In fact, one can view curvature in two ways. Intrinsic curvature is the kind of curvature that can be observed while staying in the manifold. However, it is also possible to embed a manifold in an ambient manifold and consider the curvature of the manifold in the ambient manifold. This is extrinsic curvature and it depends on the ambient space. The intrinsic curvature is an intrinsic property of the manifold and hence the most important one. However, extrinsic curvature is sometimes more intuitive. For example, a cylinder is intrinsically flat, but it is curved when it is viewed as a subset of \mathbb{R}^3 , which we are used to. We will mostly be interested in intrinsic curvature. There are multiple ways to define and characterize it. We will first define the Riemannian curvature tensor. Then the Ricci and scalar curvature can be deduced from this. We will follow [12, Chapter 7].

Parallel translation along a closed curve

We have already seen that we can transport a vector along a curve in a parallel way. One can now wonder what happens if the curve is closed. In other words, what if you come back to the point where you started? Does that yield the same vector? In \mathbb{R}^n , the answer is yes. In a more general space, however, this is not the case anymore. For example, consider the sphere. Suppose you start walking from the north pole with a stick pointing south. Once you reach the equator, walk along it for a quarter of its length. Then walk back to the north pole. If the stick is kept parallel at all times, it is still pointing to the south. This means that when you get back at the north pole, the stick has turned 90 degrees. This indicates that the sphere is curved. The reason is that differentiation in the different directions can not be exchanged.

Curvature endomorphism

More generally in \mathbb{R}^n we have $\nabla_X \nabla_Y Z = \nabla_X (Y Z^k \partial_k) = X Y Z^k \partial_k$, so

$$\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z = (XYZ^k - YXZ^k)\partial_k = ([X, Y]Z^k)\partial_k = \nabla_{[X, Y]}Z.$$

In a curved space, this does not always hold. Hence the following definition.

<u>Definition 3.33</u>. The curvature endomorphism is the map $R : \mathscr{T}(M) \times \mathscr{T}(M) \to \mathscr{T}(M) \to \mathscr{T}(M)$ given by

$$(X, Y, Z) \mapsto R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

We can define its coefficients in local coordinates by writing $R(\partial_i, \partial_j)\partial_k = R_{ijk}{}^l\partial_l$. These define a (new) map (which we will also call the curvature endomorphism) $R : \mathscr{T}(M) \times \mathscr{T}(M) \times \mathscr{T}(M) \to \mathbb{R}$ with the following action:

$$(\partial_i, \partial_j, \partial_k, \mathrm{d}x^l) \mapsto \mathrm{d}x^l(R(\partial_i, \partial_j)\partial_k) = \mathrm{d}x^l(R_{ijk}{}^l\partial_l) = R_{ijk}{}^l.$$

This means that it is indeed given by $R = R_{ijk}{}^l dx^i \otimes dx^j \otimes x^k \otimes \partial_l$ (where we interpret $\partial_l = \frac{\partial}{\partial x^l}|_p \in (T_p M)^{**}$ in each point p). In this way we turned the curvature endomorphism into a $\binom{3}{1}$ -tensor field.

Curvature tensor

Now we can give the final definition.

<u>Definition 3.34</u>. Let *R* denote the just defined tensor field. We define the *Riemannian curvature* tensor *Rm* to be *R* with the last index lowered. This means that $Rm = R_{ijkl} dx^i \otimes dx^j \otimes dx^k \otimes dx^l$ where $R_{ijkl} = g_{lm} R_{ijk}^{m}$.

Intrinsic property

We have used some tricks concerning vector spaces to end up with the covariant 4-tensor that we call the curvature tensor. However, it was worth the effort, since what we ended up with is an intrinsic property of the manifold. To make this precise, we have the following definition.

<u>Definition 3.35</u>. Let M, \tilde{M} be Riemannian manifolds with Riemannian metrics g and \tilde{g} , respectively. A map $\phi : M \to \tilde{M}$ is called an *isometry* if it is a diffeomorphism and $\phi^* \tilde{g} = g$. $\psi : M \to \tilde{M}$ is called a *local isometry* if every point in $p \in M$ has a neighbourhood U such that $\psi|_U$ is an isometry $U \to \psi(U)$. If such ψ exists, we call M and \tilde{M} *locally isometric*.

Being locally isometric thus means that the Riemannian structure is locally similar. We can now think of a property as intrinsic to a manifold if it is invariant under local isometries. This is what the next lemma says about the curvature tensor.

Lemma 3.36 ([12, Lemma 7.2]). Let M and \tilde{M} be manifolds with corresponding curvature endomorphisms and tensors R, \tilde{R}, Rm and \tilde{Rm} respectively. Suppose $\varphi : M \to \tilde{M}$ is a local isometry. Then $\varphi^* \tilde{Rm} = Rm$ and $\tilde{R}(\varphi_* X, \varphi_* Y) \varphi_* Z = \varphi_*(R(X, Y)Z)$.

Flat manifolds

We defined the curvature endomorphism as the difference between two values that should be equal in \mathbb{R}^n . This means that the curvature endomorphism is 0. This implies that the curvature tensor is also 0. Hence \mathbb{R}^n has no curvature, which is what we would like. Actually we will define flatness in this way.

<u>Definition 3.37</u>. We call a Riemannian manifold M flat if it is locally isometric to \mathbb{R}^n .

This resembles the definition of a manifold. There we require that every point has a neighbourhood that 'looks like' \mathbb{R}^n topologically. Here we require that such a neighbourhood should also resemble an open set in \mathbb{R}^n as to its Riemannian structure. Since the curvature tensor is invariant under local isometries, we directly see that any flat manifold has curvature 0. In fact, the reverse is also true.

Theorem 3.38 ([12, Thm 7.3]). Let M be a Riemannian manifold with curvature tensor Rm. Then M is flat if and only if $Rm \equiv 0$.

Ricci and scalar curvature

To have a somewhat simpler object to work with, we can derive another way to deal with curvature from the Riemannian curvature tensor.

<u>Definition 3.39</u>. The *Ricci curvature* is a covariant 2-tensor that is defined as the trace of the curvature endomorphism (considered as a $\binom{3}{1}$ -tensor) on its first and last indexes.

This means that the Ricci tensor is given by $R = R_{ij} dx^i \otimes dx^j$ with $R_{ij} = R_{kij}^k$ (where R_{ijk}^l are the coefficients of the curvature endomorphism as before). From this we can derive an even simpler notion.

<u>Definition 3.40</u>. The scalar curvature is defined as $S = g^{ij}R_{ij}$, where R_{ij} are the coefficients of the Ricci tensor. It can be obtained by raising the second index of the Ricci tensor and then taking the trace of the result.

3.3 Brownian motion

In this section we define Brownian motion on a manifold and introduce some basic operators related to it.

3.3.1 Laplace-Beltrami operator

Since the generator of Brownian motion in \mathbb{R}^n is the Laplace operator, we will define an analogue on a manifold.

Let f be a smooth function $\mathbb{R}^n \to \mathbb{R}$. Then the Laplace operator applied to f equals

$$\Delta f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial (x^i)^2} = \sum_{i=1}^{n} \frac{\partial}{\partial x^i} \frac{\partial f}{\partial x^i} = \operatorname{div}\left(\frac{\partial f}{\partial x^1}, ..., \frac{\partial f}{\partial x^n}\right) = \operatorname{div}(\nabla f).$$

This suggest that if we can define the divergence and the gradient operator on a manifold, we have a way to define the Laplace operator.

Gradient

The gradient on a manifold can be defined as follows.

<u>Definition 3.41</u>. Let f be a smooth function on M. Then the gradient of f is defined as the smooth vector field $\nabla f = df^{\sharp}$.

This means that for $v \in T_p M$: $df(v) = \langle \nabla f, v \rangle$. Note that df(v) = v(f) is the derivative of f in the direction of v. This number can be obtained as an inner product of v with the gradient vector, which is the same as in \mathbb{R}^n . To obtain a local representation, fix a point pand some coordinate system (x, U). By setting $v = \partial_i$ and writing $\nabla f = w^j \partial_j$, we can express both sides of $df(v) = \langle \nabla f, v \rangle$ in coordinates. First we see $df(v) = df(\partial_i) = \partial_i f$. Further $\langle \nabla f, \partial_i \rangle = w^j \langle \partial_j, \partial_i \rangle = w^j g_{ji} = w^j g_{ij} = (Gw)_i$. This shows that $(\partial_1 f, ..., \partial_n f)^T = Gw$, so $w = G^{-1}(\partial_1 f, ..., \partial_n f)^T$, so $w_i = g^{ij} \partial_j f$. This means that we can write

$$\nabla f = g^{ij} \partial_j f \partial_i.$$

In the Euclidean case G = I so we get the usual gradient operator back.

Divergence

In the \mathbb{R}^n case, gradient and divergence are closely related to each other. This same relation can be used as a definition of divergence on a Riemannian manifold.

Theorem 3.42 ([7, Thm 3.11]). For any smooth vector field X on a Riemannian manifold M, there is a unique smooth function $\operatorname{div}(X) : M \to \mathbb{R}$ such that for any smooth function u on M:

$$\int_{M} \operatorname{div}(X) u \, \mathrm{d}V = -\int_{M} \langle X, \nabla u \rangle \, \mathrm{d}V$$

We can also express divergence locally. For a smooth vector field X it can be shown that

$$\operatorname{div}(X) = \frac{1}{\sqrt{\det G}} \partial_i \left(\sqrt{\det G} X^i \right).$$

Note that in the Euclidean case we have det $G = \det I = 1$ on all of U, so what remains is $\operatorname{div}(X) = \partial_i X^i$, which is what we want. A closer look at the proof of [7, Thm 3.11] reveals that the factors $\sqrt{\det G}$ come from changing to the uniform measure in local coordinates and back to the volume measure.

Laplace-Beltrami operator

We can now define the analogue of the Laplace operator.

<u>Definition 3.43</u>. Let M be a Riemannian manifold. We define for smooth functions f, the Laplace-Beltrami operator Δ_M as div (∇f) .

In a coordinate system (x, U), this operator is given by:

$$\Delta_M f = \frac{1}{\sqrt{\det G}} \frac{\partial}{\partial x^j} \left(\sqrt{\det G} g^{ij} \frac{\partial f}{\partial x^i} \right).$$

According to [9, Section 2], this can be rewritten as

$$\Delta_M = g^{ij} \partial_i \partial_j - g^{ij} \Gamma^k_{ij} \partial_k.$$

It can be shown that the Laplace-Beltrami operator locally equals the following expression ([13, Section 2.3]):

$$\Delta_M f(p) = \sum_{i=1}^n \left. \frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\exp_p(tv_i)) \right|_{t=0},$$

where $v_1, ..., v_n$ is any orthogonal basis of $T_p M$. This definition resembles the expression of the Laplace operator in \mathbb{R}^n very much. Apparently we cannot just define $\Delta_M = \sum_i \partial_i^2$, because we should follow geodesics instead of straight lines in local coordinates.

3.3.2 Heat kernel

Heat equation and heat kernel

Now that we have a Laplace operator, we can formulate the heat equation on a Riemannian manifold. We say that a function $u : [0, \infty) \times M \to \mathbb{R}$ such that $(t, p) \mapsto u(t, p)$ satisfies the heat equation with initial condition f if the following holds:

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t,p) = \Delta_M u(t,p) \quad \text{for all } t > 0$$

$$\lim_{t \neq 0} u(t,\cdot) = f \text{ (locally) uniformly on } M.$$
(12)

Here Δ is applied to u as a function of p. Of course u must be smooth enough such that the derivations are well-defined. It can be shown that for each $t \geq 0$ there exists a function $p_t: M^2 \to \mathbb{R}$ from which all solutions to the heat equation with given initial conditions can be built. In particular,

$$u(t,p) = p_t * f(p) = \int_M p_t(p,q) f(q) V(\mathrm{d}q)$$

satisfies (12). The collection of functions $(p_t)_{t\geq 0}$ is called a heat kernel. [7] contains an elaborate introduction to heat kernels and the proof of their existence on Riemannian manifolds. In \mathbb{R}^n the heat kernel equals

$$p_t(x,y) = \frac{1}{(4\pi t)^{n/2}} e^{-\frac{||x-y||^2}{4t}}$$

Corresponding stochastic process

Now it is possible to define a process $W = (W_t)_{t\geq 0}$ on M such that W_t started from p has density $p_t(p, \cdot)$ (with respect to V). It can be shown that $\int_M p_t(p,q)V(dq) \leq 1$ for all t, however there is not always equality. In that case $p_t(p, \cdot)$ is the density of a sub-probability measure, and W is sub-Markov process. However, it can be shown that there is equality whenever the curvature of the space is bounded from below. As we will see in lemma 5.3 and 5.4, this is always the case in a compact manifold. So in our case we get a Markov process. Then we see that

$$u(t,p) = \int_M p_t(p,q)f(q)V(\mathrm{d}q) = \mathbb{E}_p f(W_t) = S_t f(p),$$

where S_t denotes the semigroup corresponding to W. Denote the generator by L. Now, using a basic property of semigroups, we obtain

$$\frac{\mathrm{d}u}{\mathrm{d}t}(t,\cdot) = \frac{\mathrm{d}}{\mathrm{d}t}S_t f = LS_t f = Lu(t,\cdot)$$

so $\frac{d}{dt}u(t,p) = Lu(t,p)$ for all $p \in M$ and all t > 0. Since u also satisfies (12) this heuristically suggests that $L = \Delta_M$. By a rescaling of time, the same can be done with $\frac{1}{2}\Delta_M$. We can define the corresponding process W to be Brownian motion on M.

3.3.3 Brownian motion

We define Brownian motion on M as the continuous process $W = (W_t)_{t\geq 0}$ on M that is generated by $\frac{1}{2}\Delta_M$. It was argued in the previous section why such a process exists. We will briefly state some important properties.

Distribution at time t

First of all, as we have seen above, the density of the process started from p at time t equals $p_{t/2}(p, \cdot)$, where p_t is the heat kernel. In \mathbb{R}^n , this distribution becomes

$$p_{t/2}(x,y) = \frac{1}{(4\pi(t/2))^{n/2}} e^{-\frac{||x-y||^2}{4(t/2)}} = \frac{1}{(2\pi t)^{n/2}} e^{-\frac{||x-y||^2}{2t}},$$

which is the density of the normal N(x,t) distribution. This corresponds to the definition of Brownian motion in \mathbb{R}^n . In our case $\int_M p_t(p,q)V(dq) = 1$ at all times (as noted above, since we assume compactness), so Brownian motion exists for all time. In general, it is possible that mass gets lost over time (in the case where $\int_M p_t(p,q)V(dq) < 1$ for some t). This corresponds to Brownian motion escaping to infinity. As mentioned before, this is only possible when the curvature is not bounded from below.

Core

We defined the Laplace-Beltrami operator for smooth functions, but its domain as the generator of Brownian motion might be larger than that. However, the smooth functions on M form a core for the domain. An elaboration on this can be found in [22]. Multiple times this fact will allow us to only look at smooth functions.

Dirichlet form

To find the Dirichlet form corresponding to Brownian motion, we first need to know that it is symmetric. Since the smooth functions form a core for the generator, by proposition 2.17 it suffices to check that for all smooth f, g:

$$\int_{M} \frac{1}{2} \Delta_{M} f g \mathrm{d}V = \int_{M} f \frac{1}{2} \Delta_{M} g \mathrm{d}V.$$
(13)

Now we write out the left hand side. Using the fact that ∇f is a smooth vector field on M and theorem 3.42, we obtain

$$\frac{1}{2}\int \operatorname{div}(\nabla f)g\,\mathrm{d}V = \frac{1}{2}\left(-\int_M \left\langle \nabla f, \nabla g \right\rangle \mathrm{d}V\right) = -\frac{1}{2}\int_M \left\langle \nabla f, \nabla g \right\rangle \mathrm{d}V.$$

In the same way, we can show that the right hand side of (13) equals the same expression. Therefore (13) holds and we conclude that Brownian motion is symmetric with respect to the volume measure.

To find an expression for the Dirichlet form, we can use the relation with the generator. Since the smooth functions are in the domain of the generator, they are also in the domain of the Dirichlet form. We see for smooth functions f and g:

$$\mathcal{E}(f,g) = \left(-\frac{1}{2}\Delta_M f,g\right) = -\frac{1}{2}\int \operatorname{div}(\nabla f)g \,\mathrm{d}V = \frac{1}{2}\int_M \langle \nabla f, \nabla g \rangle \,\mathrm{d}V.$$

We thus conclude that for smooth f, g the Dirichlet form corresponding to Brownian motion equals

$$\mathcal{E}(f,g) = \frac{1}{2} \int_M \langle \nabla f, \nabla g \rangle \,\mathrm{d}V.$$

Now fix a coordinate system (x, U). ∇ is locally given by $\nabla f = g^{ik} \partial_i f \partial_k$, so we see

$$\begin{split} \langle \nabla f, \nabla h \rangle &= \left\langle g^{ik} \partial_i f \partial_k, g^{jl} \partial_j h \partial_l \right\rangle = g^{ik} g^{jl} \partial_i f \partial_j h \left\langle \partial_k, \partial_l \right\rangle \\ &= g^{ik} g^{jl} \partial_i f \partial_j h g_{kl} = \sum_{i,j=1}^n \partial_i f \partial_j h \sum_{l=1}^n g^{jl} \sum_{k=1}^n g^{ik} g_{kl} \\ &= \sum_{i,j=1}^n \partial_i f \partial_j h \sum_{l=1}^n g^{jl} \delta_l^i = \partial_i f \partial_j h g^{ji} = \partial_i f \partial_j h g^{ij}. \end{split}$$

In the case of \mathbb{R}^n , this reduces to $\sum_{i=1}^n \partial_i f \partial_i h$, so in that case

$$\mathcal{E}(f,f) = \frac{1}{2} \int \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x^{i}}\right)^{2} \mathrm{d}\lambda,$$

which is what we want.

4 The invariance principle on a Riemannian manifold

Let M be a Riemannian manifold with dimension n. We assume that M is compact. Then we know that M is complete and hence geodesically complete. Moreover, we assume that M is connected. In section 4.1, we will first define a simple random walk on M. Then we show that the generators of these random walks converge to the generator of Brownian motion when the step size decreases and the step rate increases. This implies that the corresponding processes converge in distribution in the path space. In section 4.2 we first discuss the existence and expression of the corresponding Dirichlet forms. Then we show convergence (in a similar way to the convergence of the generators) and we make a remark on what the convergence of Dirichlet forms says about the convergence of processes. While showing convergence of the generators (and of the Dirichlet forms) we discover which assumptions should be made about the stepping distributions for this result to hold. Section 4.3 is devoted to finding out which distributions satisfy these assumptions.

4.1 Convergence of the generators

The process

Suppose that M is equipped with a collection of measures $\{\mu_p, p \in M\}$ such that μ_p is a measure on T_pM . For a vector $\xi \in T_pM$ we denote the geodesic through p with tangent vector ξ at pby $p(\cdot,\xi)$. We define a random walk $X^N = \{X_t^N, t \ge 0\}$ on M starting from $q \in M$ as the process that has $X_0 = q$ and that jumps from p after an exponential time (with parameter 1) to $p(1/N, \eta)$ where $\eta \in T_pM$ is drawn according to a probability distribution μ_p . So at each point, the process waits an exponential time, then picks a tangent vector η according to μ_p and then walks for time $\frac{1}{N}$ along the geodesic through p in the direction of η . This situation is sketched in figure 3. We assume that the exponential waiting times and the random tangent vectors are all independent. As we move through this section, we will see which restrictions the measures μ_p should satisfy. In particular, we will show (in section 4.3) that we can take μ_p to be for instance the uniform distribution on $\sqrt{n}S_pM$. Here S_pM denotes the set of unit tangent vectors at p and for a > 0 aS_pM denotes S_pM multiplied by a (all vectors with norm a). The associated semigroup on C(M) is

$$S_t^N f(p) = \mathbb{E}_p f(X_t^N).$$

This semigroup is generated by

$$L_N f(p) = \int_{T_p M} f(p(1/N, \eta)) - f(p)\mu_p(\mathrm{d}\eta)$$

This is where we need independence of the waiting times and the random vectors. Note that this is just an analogous expression to the generator of a random walk in section 2.1. The domain is the set of all continuous functions on M.

The \mathbb{R}^n case

In the introduction we already described the result in \mathbb{R}^n , but we will repeat it here in a way that is a bit more technical to see that the process described above generalizes it. We have seen that in \mathbb{R}^n the exponential map is just addition if we identify $T_p\mathbb{R}^n$ with \mathbb{R}^n itself. So in that case from a point p the process moves to $p + \frac{1}{N}\eta$ where η is chosen from $T_p\mathbb{R}^n = \mathbb{R}^n$ randomly.



Figure 3: Left: geodesic random walk on a sphere. Right: Brownian motion on a sphere (source: https://en.wikipedia.org/wiki/Brownian_motion).

This means that the discrete time jumping process when jumping as described above, can be denoted by $S_m^N = \sum_{i=1}^m \frac{1}{N} \eta_i = \frac{1}{N} \sum_{i=1}^m \eta_i$ where η_j is drawn from $T_{S_{j-1}} \mathbb{R}^n = \mathbb{R}^n$ according to some distribution. Now let $\{N_t, t \ge 0\}$ be a Poisson process with rate one and define $X_t^N = S_{N_t}$. Then X makes the same jumps as S, but after independent exponential times. We see that $X^N = \{X_t^N, t \ge 0\}$ satisfies the description above. Now the invariance principle tells us that under some restrictions on the jumping distributions $X_{tN^2}^N \to B_t$ in distribution as N goes to infinity. Here B is Brownian motion. We will try to mimic this in the more general setting of a manifold.

Aim

We would like to show that a rescaled version of the random walk described above converges to Brownian motion as N goes to infinity (so that the left picture of figure 3 converges to the right one). In \mathbb{R}^n we had to scale time by N^2 if space is scaled by N. This turns out to be the right scaling in the general case as well.

Theorem 2.13 (Trotter-Kurtz) says that for convergence in distribution in $D([0,\infty), M)$ (the space of cadlag maps $[0,\infty) \to M$) it suffices to show convergence of the generators on a core for the limit process. Since $C^{\infty}(M)$ is a core for $\frac{1}{2}\Delta_M$, it suffices to consider smooth functions. Considering all of the above, our aim is to show that for any smooth f we have $N^2 L_N f \to \frac{1}{2}\Delta_M f$ uniformly on M.

Choosing the right charts

Let f be a fixed smooth function from now on. Since we want the convergence $N^2 L_N f \to \frac{1}{2} \Delta_M f$ to be uniform on M, we cannot just consider this problem pointwise. To deal with this, we will choose specific coordinate charts.

Let (as before) ρ denote the original metric of the manifold and let d denote the metric that is induced by the Riemannian metric. Recall that these metrics induce the same topology. This means that we do not cause confusion when we speak about open and closed sets, continuous maps and compactness without explicitly mentioning the metric. For each $p \in M$, let (x_p, U_p) be a coordinate chart for M around p. U_p is open with respect to ρ and hence with respect to d. This means that there is some $\epsilon_p > 0$ such that $G_p := \overline{B_d(p, \epsilon_p)} \subset U_p$. Now define $O_p = B_d(p, \epsilon/2)$. Since M is compact, we can find $p_1, ..., p_m$ such that $M \subset \cup_i O_{p_i}$. We have the following easy statement. <u>Lemma 4.1</u>. Let $(g_k)_{k=1}^{\infty}$ and g be functions $M \to \mathbb{R}$. If $g_k \to g$ uniformly on each O_{p_i} , then $g_k \to g$ uniformly on M.

Proof. Let $\epsilon > 0$. For each *i* there is an N_i such that for all $k \ge N_i : \sup_{O_{p_i}} |g_k(q) - g(q)| < \epsilon$. Set $N = \max_{1 \le i \le m} N_i$ and let $q \in M$. Then there is a *j* such that $q \in O_{p_j}$. Now for all $k \ge N$, we see $k \ge N_j$, so $|g_k(q) - g(q)| \le \sup_{O_{p_i}} |g_k(s) - g(s)| < \epsilon$. This shows that $\sup_M |g_k(q) - g(q)| \le \epsilon$. Hence $g_k \to g$ uniformly on M.

Now let $j \in \{1, ..., m\}$ be fixed. Call $O := O_{p_j}$, $\epsilon := \epsilon_{p_j}$, $x := x_{p_j}$, $G := G_{p_j}$ and $U := U_{p_j}$ (this situation is shown in figure 4). Because of the lemma, it suffices to show that $N^2 L_N f \to \frac{1}{2} \Delta_M f$ uniformly on O.

Technical considerations

To obtain good estimations later, we will need that $p(s,\eta)$ is still in our coordinate system (x, U) and even in the set G when $|s| \leq \frac{1}{N}$ for N large enough. Since the convergence must be uniform, how large N must be can not depend on the point p. The following lemma tells us how to choose such N.

<u>Lemma 4.2</u>. Suppose that for every $p \in M$ there exists a K_p such that $||\eta|| < K_p$ for all $\eta \in \operatorname{supp}\mu_p$. Moreover, suppose that $K = \sup_M K_p < \infty$. Choose $N_{\epsilon} \in \mathbb{N}$ such that $\frac{1}{N_{\epsilon}} < \frac{\epsilon}{2K}$. Then for all $p \in O$ and $N \ge N_{\epsilon}$ we see

$$\forall |s| \le \frac{1}{N} : p(s,\eta) \in G.$$

Proof. Let $N \ge N_{\epsilon}$ and let $p \in O$. The situation of the proof is visually represented in figure 4. Fix $s \in (-\frac{1}{N}, \frac{1}{N})$. Without loss of generality assume s > 0. Note that the speed of the geodesic $p(\cdot, \eta)$ equals $||\eta||$, so at time s, it has traveled a distance $s||\eta||$ from p. This means that there is a path of length $s||\eta||$ from $p(s,\eta)$ to p, so $d(p(s,\eta),p) \le s||\eta|| \le \frac{1}{N}K \le \frac{1}{N_{\epsilon}}K < \epsilon/2$. Since $p \in O$, we know $d(p,p_j) < \epsilon/2$. Now the triangle inequality shows that $d(p_j,p(s,\eta)) \le d(p_j,p) + d(p,p(s,\eta)) < \epsilon/2 + \epsilon/2 = \epsilon$. This implies that $p(s,\eta) \in B_d(p_j,\epsilon) \subset G$.



Figure 4: The chart (x, U) with closed ball G and open ball O around p_j . As is shown in lemma 4.2, $p^{\eta} = p(t, \eta)$ does not leave the ball around p with radius $\epsilon/2$, as long as $|t| \leq 1/N$ for $N \geq N_{\epsilon}$. The importance for uniformity is that it does not matter where we choose p (in O).

Note that to do this, we must assume that the support of μ_p is compact for each p and in some sense uniformly in p (so K as above exists). If not, the geodesics can go arbitrarily fast, so they generally do not stay close enough to the starting point.

Assumption 1: μ_p has compact support for each $p \in M$. Moreover

$$\sup_{p \in M} \sup_{\eta \in \mathrm{supp}\mu_p} ||\eta|| < \infty.$$

We know now that K as in lemma 4.2 exists. We fix an N_{ϵ} as in the lemma and take N larger than N_{ϵ} .

Taylor expansion

Now fix $p \in O$ and $\eta \in T_p M$. Write p^{η} for the map $\mathbb{R} \to M$ that takes t to $p(t,\eta)$. We can locally write $f \circ p^{\eta} = (f \circ x^{-1}) \circ (x \circ p^{\eta})$, which is a composition of smooth maps. This means that $f \circ p^{\eta}$ is just a smooth map $\mathbb{R} \to \mathbb{R}$, so we can use a Taylor expansion and obtain

$$f(p(1/N,\eta)) = f(p) + \frac{1}{N} \frac{\mathrm{d}(f \circ p^{\eta})}{\mathrm{d}t}(0) + \frac{1}{2N^2} \frac{\mathrm{d}^2(f \circ p^{\eta})}{\mathrm{d}^2t}(0) + \frac{1}{6N^3} \frac{\mathrm{d}^3(f \circ p^{\eta})}{\mathrm{d}^3t}(t_{N,\eta,p}),$$

where $t_{N,\eta,p} \in (0, 1/N)$ is a number depending on N, η and p. This gives us

$$N^{2}L_{N}f(p) = N^{2}\int_{M_{p}} f(p(1/N,\eta)) - f(p)\mu_{p}(\mathrm{d}\eta)$$

$$= N^{2}\int \frac{1}{N}\frac{\mathrm{d}(f\circ p^{\eta})}{\mathrm{d}t}(0) + \frac{1}{2N^{2}}\frac{\mathrm{d}^{2}(f\circ p^{\eta})}{\mathrm{d}^{2}t}(0) + \frac{1}{6N^{3}}\frac{\mathrm{d}^{3}(f\circ p^{\eta})}{\mathrm{d}^{3}t}(t_{N,\eta,p})\mu_{p}(\mathrm{d}\eta)$$

$$= N\int \frac{\mathrm{d}(f\circ p^{\eta})}{\mathrm{d}t}(0)\mu_{p}(\mathrm{d}\eta) + \frac{1}{2}\int \frac{\mathrm{d}^{2}(f\circ p^{\eta})}{\mathrm{d}t^{2}}(0)\mu_{p}(\mathrm{d}\eta) + \frac{1}{6N}\int \frac{\mathrm{d}^{3}(f\circ p^{\eta})}{\mathrm{d}t^{3}}(t_{N,\eta,p})\mu_{p}(\mathrm{d}\eta).(14)$$

We will examine these terms separately.

The first term

Recall that $p \in O$ and that O is contained in a coordinate chart (x, U). Since $N \ge N_{\epsilon}$, lemma 4.2 guarantees us that $p(s, \eta)$ stays in the coordinate chart for $|s| < \frac{1}{N}$. Writing $\eta = \sum_{i=1}^{n} \eta^{i} \frac{\partial}{\partial x^{i}}|_{p}$, we see for $|s| < \frac{1}{N}$:

$$\begin{aligned} \frac{\mathrm{d}(f \circ p^{\eta})}{\mathrm{d}t}(s) &= \frac{\mathrm{d}}{\mathrm{d}t} [(f \circ x^{-1}) \circ (x \circ p^{\eta})](s) \\ &= \sum_{i=1}^{n} D_i (f \circ x^{-1}) (x(p^{\eta}(s)) \frac{\mathrm{d}(x^i \circ p^{\eta})}{\mathrm{d}t}(s) \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} (p^{\eta}(s)) \frac{\mathrm{d}(x^i \circ p^{\eta})}{\mathrm{d}t}(s). \end{aligned}$$

Now setting s = 0, this becomes:

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x^{i}}(p)\eta^{i} = \sum_{i=1}^{n} \eta^{i} \frac{\partial}{\partial x^{i}}|_{p}f = \eta(f),$$

since $p^{\eta}(0) = p(0, \eta) = p$ and the tangent vector to the geodesic $p(\cdot, \eta)$ at 0 is η (so the *i*th coordinate with respect x is just η^i). Now the first term of (14) becomes:

$$N\int \eta(f)\mu_p(\mathrm{d}\eta) = N\int \sum_{i=1}^n \eta^i \frac{\partial}{\partial x^i}|_p f\mu_p(\mathrm{d}\eta) = N\sum_{i=1}^n \frac{\partial}{\partial x^i}|_p f\int \eta^i \mu_p(\mathrm{d}\eta).$$

This shows that the first term of (14) vanishes if we make the following assumption. **Assumption 2:** For every coordinate chart around any $p \in M$: $\int \eta^i \mu_p(d\eta) = 0$ for each *i*.

The second term

Now we want to show that the remaining term equals $\frac{1}{2}\Delta_M f(p)$. Similarly to above we see for

 $|s| < \frac{1}{N}$ (leaving out the arguments to keep things clear):

$$\begin{aligned} \frac{\mathrm{d}^2(f \circ p^\eta)}{\mathrm{d}t^2} &= \frac{\mathrm{d}}{\mathrm{d}t} \sum_{i=1}^n \frac{\partial f}{\partial x^i} \frac{\mathrm{d}(x^i \circ p^\eta)}{\mathrm{d}t} \\ &= \sum_{i=1}^n \left\{ \left(\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial f}{\partial x^i} \right) \frac{\mathrm{d}(x^i \circ p^\eta)}{\mathrm{d}t} + \frac{\partial f}{\partial x^i} \left(\frac{\mathrm{d}}{\mathrm{d}t} \frac{\mathrm{d}(x^i \circ p^\eta)}{\mathrm{d}t} \right) \right\} \\ &= \sum_{i=1}^n \left\{ \sum_{j=1}^n \frac{\partial^2 f}{\partial x^j \partial x^i} \frac{\mathrm{d}(x^j \circ p^\eta)}{\mathrm{d}t} \frac{\mathrm{d}(x^i \circ p^\eta)}{\mathrm{d}t} + \frac{\partial f}{\partial x^i} \frac{\mathrm{d}^2(x^i \circ p^\eta)}{\mathrm{d}t^2} \right\}.\end{aligned}$$

Since p^{η} is a geodesic, we know that it satisfies the geodesic equations. This shows that for each i = 1, .., n we have

$$\frac{\mathrm{d}^2(x^i \circ p^\eta)}{\mathrm{d}t^2} + \sum_{k,l=1}^n \Gamma^i_{kl} \frac{\mathrm{d}(x^k \circ p^\eta)}{\mathrm{d}t} \frac{\mathrm{d}(x^l \circ p^\eta)}{\mathrm{d}t} = 0.$$

Using this yields the following expression for the second derivative:

$$\sum_{i=1}^{n} \left\{ \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x^j \partial x^i} \frac{\mathrm{d}(x^j \circ p^\eta)}{\mathrm{d}t} \frac{\mathrm{d}(x^i \circ p^\eta)}{\mathrm{d}t} - \frac{\partial f}{\partial x^i} \sum_{k,l=1}^{n} \Gamma_{kl}^i \frac{\mathrm{d}(x^k \circ p^\eta)}{\mathrm{d}t} \frac{\mathrm{d}(x^l \circ p^\eta)}{\mathrm{d}t} \right\},$$

 \mathbf{SO}

$$\frac{\mathrm{d}^2(f \circ p^\eta)}{\mathrm{d}t^2}(0) = \sum_{i=1}^n \left\{ \sum_{j=1}^n \frac{\partial^2 f}{\partial x^j \partial x^i}(p) \eta^j \eta^i - \frac{\partial f}{\partial x^i}(p) \sum_{k,l=1}^n \Gamma^i_{kl}(p) \eta^k \eta^l \right\}.$$

Using linearity of the integral, we obtain the following expression for the second term of (14):

$$\frac{1}{2}\sum_{i=1}^{n}\left\{\sum_{j=1}^{n}\frac{\partial^{2}f}{\partial x^{i}\partial x^{j}}(p)\int\eta^{i}\eta^{j}\mu_{p}(\mathrm{d}\eta)-\frac{\partial f}{\partial x^{i}}(p)\sum_{k,l=1}^{n}\Gamma_{kl}^{i}(p)\int\eta^{k}\eta^{l}\mu_{p}(\mathrm{d}\eta)\right\}$$

Note that we also changed the order of the derivatives of f, this can be done since f is smooth. Now we want the term above to equal

$$\frac{1}{2}\Delta_M f(p) = \frac{1}{2} \left\{ g^{ij} \frac{\partial^2 f}{\partial x^i x^j} - g^{kl} \Gamma^i_{kl} \frac{\partial f}{\partial x^i} \right\}$$
$$= \frac{1}{2} \sum_{i=1}^n \left\{ \sum_{j=1}^n \frac{\partial^2 f}{\partial x^i \partial x^j}(p) g^{ij}(p) - \frac{\partial f}{\partial x^i}(p) \sum_{k,l=1}^n \Gamma^i_{kl}(p) g^{kl}(p) \right\}.$$

This is true if we require the following.

Assumption 3: For any coordinate chart around any $p \in M$: $\int_{M_p} \eta^i \eta^j \mu_p(d\eta) = g^{ij}(p)$ for all i, j.

The rest term

If the last term goes to 0 uniformly on O, we have the result. Let N still be larger than N_{ϵ} .

$$\begin{aligned} \left| \frac{1}{6N} \int \frac{\mathrm{d}^3(f \circ p^{\eta})}{\mathrm{d}t^3}(t_{N,\eta,p}) \mu_p(\mathrm{d}\eta) \right| &\leq \frac{1}{6N} \int \left| \frac{\mathrm{d}^3(f \circ p^{\eta})}{\mathrm{d}t^3}(t_{N,\eta,p}) \right| \mu_p(\mathrm{d}\eta) \\ &\leq \frac{1}{6N} \sup_{\eta \in \mathrm{supp}\mu_p} \left| \frac{\mathrm{d}^3(f \circ p^{\eta})}{\mathrm{d}t^3}(t_{N,\eta,p}) \right| \end{aligned}$$

Note that we used here that μ_p is a probability measure. We know that $t_{N,\eta,p} \in [0, 1/N] \subset [0, 1/N_{\epsilon}]$. This means that the above is smaller than:

$$\frac{1}{6N} \sup_{\eta \in \mathrm{supp}\mu_p} \sup_{t \in [0, 1/N_{\epsilon}]} \left| \frac{\mathrm{d}^3(f \circ p^{\eta})}{\mathrm{d}t^3}(t) \right| \le \frac{1}{6N} \sup_{\eta : ||\eta|| \le K} \sup_{t \in [0, 1/N_{\epsilon}]} \left| \frac{\mathrm{d}^3(f \circ p^{\eta})}{\mathrm{d}t^3}(t) \right|.$$

Because of the 1/N in front of the equation, we only need to know that the rest is uniformly bounded to obtain uniform convergence. It thus suffices to show that $\frac{d^3(f \circ p^{\eta})}{dt^3}(t)$ is bounded as a function of η with $||\eta|| < K$ and $t \in [0, 1/N_{\epsilon}]$. Lemma 4.2 shows that $p(t, \eta)$ stays in G for all such η and t. We will use this fact multiple times.

We first express $\frac{d^3(f \circ p^{\eta})}{dt^3}$ in local coordinates for $|t| \le 1/N$.

$$\frac{\mathrm{d}^3(f\circ p^\eta)}{\mathrm{d}t^3} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}^2(f\circ p^\eta)}{\mathrm{d}t^2} = \frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^n \left\{\sum_{j=1}^n \frac{\partial^2 f}{\partial x^j \partial x^i}\frac{\mathrm{d}(x^j\circ p^\eta)}{\mathrm{d}t}\frac{\mathrm{d}(x^i\circ p^\eta)}{\mathrm{d}t} + \frac{\partial f}{\partial x^i}\frac{\mathrm{d}^2(x^i\circ p^\eta)}{\mathrm{d}t^2}\right\}.$$
(15)

To make notation more compact, we introduce the following notation (and f_i , f_{ijk} analogously):

$$f_{ij} := \frac{\partial^2 f}{\partial x^j \partial x^i}, \qquad \qquad p_k^i := \frac{\mathrm{d}^k (x^i \circ p^\eta)}{\mathrm{d} t^k}.$$

Combining this with Einstein summation, we can write (15) as

$$\frac{\mathrm{d}}{\mathrm{d}t}(f_{ij}p_1^ip_1^j + f_ip_2^i) = (f_{ijk}p_1^k)p_1^ip_1^j + f_{ij}(p_1^ip_2^j + p_2^ip_1^j) + (f_{ij}p_1^j)p_2^i + f_ip_3^i$$

$$= f_{ijk}p_1^kp_1^ip_1^j + f_{ij}(p_1^ip_2^j + 2p_2^ip_1^j) + f_ip_3^i.$$

Now, as before, we can deal with second derivatives of geodesics using the geodesic equations:

$$p_2^i = -\Gamma_{rs}^i p_1^r p_1^s.$$

We can also calculate the third derivative:

$$p_3^i = \frac{\mathrm{d}}{\mathrm{d}t} p_2^i = \frac{\mathrm{d}}{\mathrm{d}t} (-\Gamma_{rs}^i p_1^r p_1^s) = -\left(\frac{\mathrm{d}}{\mathrm{d}t} \Gamma_{rs}^i\right) p_1^r p_1^s - \Gamma_{rs}^i (p_1^r p_2^s + p_2^r p_1^s).$$

This shows us that $\frac{d^3(f \circ p^{\eta})}{dt^3}$ is a combination of products and sums of the following types of expressions: f_i , f_{ij} , f_{ijk} , p_1^i , Γ_{rs}^i and $\frac{d}{dt}\Gamma_{rs}^i$. If we can bound all of these on the right domains (independent of p and η), we are done.

Bounding f_i , f_{ij} and f_{ijk}

First of all, note that f is a smooth function on U. Further, ∂_i defines smooth vector field on U. Since $f_i = \frac{\partial f}{\partial x^i}$ is obtained by applying ∂_i on U to f, it is a smooth function on U. Continuing in this way, we see that f_{ij} and f_{ijk} are also smooth functions on U. In particular, they are smooth functions on G (since it is a subset of U). G is a closed subset of the compact M and is hence compact itself. This implies that f_i , f_{ij} and f_{ijk} are (for each choice of i, j, k) bounded on G. Since we evaluate these functions in the points $p(s, \eta)$ for $0 \le s \le 1/N$, $N \ge N_{\epsilon}$ and $||\mu|| \le K$, our discussion above shows that we only evaluate them in points of G. This means that we have found bounds for f_i , f_{ij} and f_{ijk} .

Bounding p_1^i

We start with a technical lemma.

<u>Lemma 4.3</u>. Let $q \in M$ and let (y, V) be a coordinate chart around q. Let $v \in T_q M$ and write $v = v^i \partial_i$. Then $|v^i| \leq \sqrt{g^{ii}(q)} ||v||$.

Proof. Fix some $1 \le i \le n$. We see in the tangent space at q:

$$\left\langle v, g^{ij}\partial_j \right\rangle = \left\langle v^k \partial_k, g^{ij}\partial_j \right\rangle = v^k g^{ij} g_{kj} = v^k \delta^i_k = v^i$$

Further,

$$||g^{ij}\partial_j||^2 = \left\langle g^{ij}\partial_j, g^{ik}\partial_k \right\rangle = g^{ij}g^{ik}g_{jk} = g^{ij}\delta^i_j = g^{ii}.$$

Using the relations above and the Cauchy-Schwarz inequality, we obtain:

$$|v^i| = |\langle v, g^{ij}\partial_j \rangle| \le ||v|| \cdot ||g^{ij}\partial_j|| = \sqrt{g^{ii}}||v||.$$

Now we can use this to show the following.

Lemma 4.4.
$$|p_1^i(t)| = \left| \frac{\mathrm{d}(x^i \circ p^\eta)}{\mathrm{d}t}(t) \right| \le \sqrt{g^{ii}(p(t,\eta))} ||\eta||.$$

Proof. The first equation is just a change of notation. Using the definitions from section A.2.3, we see

$$\frac{\mathrm{d}(x^i \circ p^\eta)}{\mathrm{d}t} = \left(p_*^\eta \frac{\mathrm{d}}{\mathrm{d}t}\right)(x^i) = \frac{\mathrm{d}p^\eta}{\mathrm{d}t}(x^i) = \left(\frac{\mathrm{d}p^\eta}{\mathrm{d}t}\right)^i.$$

This means that $\frac{d(x^i \circ p^{\eta})}{dt}$ is just the *i*th coordinate with respect to (x, U) of the tangent vector to p^{η} at time t so at the point $p(t, \eta) \in M$. Using lemma 4.3, we see

$$\left|\frac{\mathrm{d}(x^{i}\circ p^{\eta})}{\mathrm{d}t}(t)\right| \leq \sqrt{g^{ii}(p(t,\eta))} \left|\left|\frac{\mathrm{d}p^{\eta}}{\mathrm{d}t}\right|\right|.$$
(16)

Since p^{η} is a geodesic, it has constant speed. Its speed at p is $||\eta||$, so this must be its speed anywhere else along the trajectory. Hence $||\frac{dp^{\eta}}{dt}|| = ||\eta||$. Inserting this in (16) yields the result.

We can now easily obtain a bound for p_1^i . For $0 \le t \le 1/N$ and $||\eta|| \le K$, we know $p(t,\eta)$ stays in G. g^{ii} is a smooth and hence continuous function on U, so it is bounded on G (since G is compact). This means that $\sqrt{g^{ii}(p(t,\eta))}$ is bounded by some K^i for $||\eta|| \le K$ and $0 \le t \le 1/N$. Now we see $|p_1^i| \le \sqrt{g^{ii}(p(t,\eta))} \left| \left| \frac{\mathrm{d}p^{\eta}}{\mathrm{d}t} \right| \right| \le K^i K$.

Bounding Γ_{rs}^i and $\frac{\mathrm{d}}{\mathrm{d}t}\Gamma_{rs}^i$

Each g_{ij} is a smooth function on U. This means that $\frac{\partial g_{ij}}{\partial x^k}$ is a smooth function on U. This implies that Γ_{rs}^i is just combination of products and sums of smooth functions, so it is smooth itself. Now, as before, Γ_{rs}^i is bounded on G. Since we only evaluate it in $p(t,\eta)$ with $0 \le t \le 1/N$ and $||\eta|| \le K$, we only evaluate it in G, so we have bounded Γ_{rs}^i . Now $\frac{d}{dt}\Gamma_{rs}^i$ can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\Gamma_{rs}^{i} = \frac{\partial\Gamma_{rs}^{i}}{\partial x^{j}}\frac{\mathrm{d}(x^{j}\circ p^{\eta})}{\mathrm{d}t} = (\Gamma_{rs}^{i})_{j}p_{1}^{j},$$

with notation as above. Since Γ_{rs}^i is smooth function $U \to \mathbb{R}$, this expression can be bounded in exactly the same way as expressions like $f_j p_1^j$ above.

Conclusion

We have shown that $N^2 L_N f \to \frac{1}{2} \Delta_M f$ uniformly for smooth f if:

- $\sup_{p \in M} \sup_{\eta \in \operatorname{supp} \mu_p} ||\eta|| < \infty$
- $\int \eta^i \mu_p(\mathrm{d}\eta) = 0$ and $\int \eta^i \eta^j \mu_p(\mathrm{d}\eta) = g^{ij}(p)$ in each coordinate system around p

Note that if we denote the random variable corresponding to μ_p by ζ_p , the second requirement is that (in any coordinate system) $\mathbb{E}\zeta_p^i = 0$ and $\operatorname{Cov}(\zeta_p^i, \zeta_p^j) = g^{ij}(p)$. This shows that the mean vector m of ζ_p satisfies m = 0 and the covariance matrix Σ satisfies $\Sigma = (g^{ij})(p)$. In \mathbb{R}^n , this simplifies to $\mathbb{E}\zeta_p^i = 0$ and $\operatorname{Cov}(\zeta_p^i, \zeta_p^j) = \delta_j^i$. This is satisfied for instance when μ_p is the uniform distribution on a ball with the right radius in \mathbb{R}^n . Section 4.3 deals with the question which measures satisfy the restrictions above. Some examples will be given at the end of that section as well.

4.2 Convergence of the Dirichlet forms

Reversibility of the processes

To use the Dirichlet forms, we first need to know that the underlying process is reversible (with respect to the volume measure). This means that we have to show that for all f, g in a core of the generator $(L_N f, g) = (f, L_N g)$, i.e.:

$$\int_{M} \int_{T_{pM}} \left(f(p(1/N,\eta)) - f(p) \right) \mu_{p}(\mathrm{d}\eta) g(p) \mathrm{d}V = \int_{M} \int_{T_{pM}} \left(g(p(1/N,\eta)) - g(p) \right) \mu_{p}(\mathrm{d}\eta) f(p) \mathrm{d}V.$$

Canceling common terms, we obtain that we need the following identity:

$$\int_M \int_{T_pM} f(p(1/N,\eta))\mu_p(\mathrm{d}\eta)g(p)\mathrm{d}V = \int_M \int_{T_pM} g(p(1/N,\eta))\mu_p(\mathrm{d}\eta)f(p)\mathrm{d}V.$$

If we consider \mathbb{R}^n with $V = \lambda$ the Lebesgue measure, this can be shown by manipulating the left hand side in the following way. First we identify $T_x \mathbb{R}^n$ with \mathbb{R}^n and choose a jumping distribution that does not depend on p. Then the order of integration is switched, to obtain (in \mathbb{R}^n notation, writing the exponential map as summation):

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x+y)\mu(\mathrm{d} y)g(x)\mathrm{d} \lambda = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x+y)g(x)\mathrm{d} \lambda \mu(\mathrm{d} y).$$

Now we transform the inner integral by translating over y and we use the translation invariance of the Lebesgue measure. Then we change the integrals back, so we obtain:

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x)g(x-y) d\lambda \mu(dy) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x)g(x-y)\mu(dy) d\lambda.$$

Now we need that μ is symmetric in the sense that we can transform -y to +y without changing μ . Doing that, we get:

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x) g(x+y) \mu(\mathrm{d}y) \mathrm{d}\lambda = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} g(x+y) \mu(\mathrm{d}y) f(x) \mathrm{d}\lambda,$$

which is what was to be shown.

This seems like a natural proof, so our hope would be to copy it in the manifold case. However, we come across several difficulties.

First of all, μ_p is a measure on T_pM . This means that the inner integral integrates over a different vector space for every p. It is therefore not so easy to speak of letting μ not depend on p, let alone to interchange the integrals.

The next difficulty is that we cannot really perform translation of the integral, since one cannot just add a fixed vector to some point. The only sensible way would be to use the exponential map, but then the question is how to define 'the same vector' everywhere (so we 'translate' in the same direction). Of course all of these tangent space are isomorphic as vector spaces, but there are many possible isomorphisms. Even if this can be done, the question is over which vector we translate. It must give some sort of inverse of following the exponential map.

The last problem is that we need the volume measure to be translation invariant. It is not so easy to see what this should mean in a manifold, since it is not clear what we mean with translation.

Example

Despite all of these problems, it is possible to have reversibility in non-Euclidean spaces. Certain groups (for example the group of $n \times n$ -matrices with determinant 1 or the group of rotations in \mathbb{R}^n) are manifolds that have the concepts that we need above: a jumping measure that does

not depend on the point, translation and a translation invariant measure (the Haar measure). For such groups, calculations analogous to the calculations for \mathbb{R}^n above can be carried out.

Corresponding Dirichlet forms

If the underlying process is reversible, we need to find out what the Dirichlet form looks like. To do that, we can use the relation $\mathcal{E}(f,g) = (-Lf,g)$, which yields

$$\mathcal{E}_N(f,g) = -\int_M \int_{T_pM} (f(p(1/N,\eta)) - f(p))g(p)\mu_p(\mathrm{d}\eta)V(\mathrm{d}p).$$

In the case of \mathbb{R}^n , it can be shown that this equals

$$-\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}(f(x+y)-f(x))g(x)\mu(\mathrm{d}y)\lambda(\mathrm{d}x) = \frac{1}{2}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}(f(x+y)-f(x))(g(x+y)-g(x))\mu(\mathrm{d}y)\lambda(\mathrm{d}x).$$

To show the last equality, we need the same kind of properties as for showing reversibility above, so it is not directly clear if the same holds in the manifold case. However, since the latter expression is the most natural one, we choose to use the analogue in the manifold case. Therefore we define:

$$\mathcal{E}_{N}(f,g) := \frac{1}{2} \int_{M} \int_{T_{p}M} (f(p(1/N,\eta)) - f(p))(g(p(1/N,\eta)) - g(p))\mu_{p}(\mathrm{d}\eta)V(\mathrm{d}p).$$

We will call it a Dirichlet form, although we should keep in mind that we have not shown the last identity and that we need reversibility for it to truly be a Dirichlet form. As we said before, the reason to use this expression is that it seems the most natural expression for a Dirichlet form in this context. It will turn out that we get the convergence result that we want. Moreover, we will need the calculations in the general (non-reversible) case in section 5.

\mathbf{Aim}

We will now analyze the same situation with the Dirichlet forms as defined above. We assume assumption 1, 2 and 3 of section 4.1. Recall that the Dirichlet form of Brownian motion is

$$\mathcal{E}(f,g) = \frac{1}{2} \int_M \langle \nabla f, \nabla g \rangle \,\mathrm{d}V.$$

We will leave out the factor $\frac{1}{2}$ of both \mathcal{E}_N and \mathcal{E} , so we do not need to write it every time. This obviously does not change the calculations.

We would like to show that $N^2 \mathcal{E}_N(f,g) \to \mathcal{E}(f,g)$ for smooth f,g. Note that M is compact, so V is a finite measure, so all smooth functions are quadratically integrable. Moreover, whenever \mathcal{E}_N is actually a Dirichlet form, its domain contains all the smooth functions, since the domain of the corresponding generator contains all those functions. Because of a polarization argument, it suffices to show the result for f = g, since then

$$N^{2}\mathcal{E}_{N}(f,g) = \frac{1}{2} \left(N^{2}\mathcal{E}_{N}(f+g,f+g) - N^{2}\mathcal{E}_{N}(f,f) - N^{2}\mathcal{E}_{N}(g,g) \right)$$

$$\rightarrow \frac{1}{2} \left(\mathcal{E}(f+g,f+g) - \mathcal{E}(f,f) - \mathcal{E}(g,g) \right) = \mathcal{E}(f,g).$$

This means that we would like to prove that:

$$N^2 \int_M \int_{T_p M} (f(p(1/N,\eta)) - f(p))^2 \mu_p(\mathrm{d}\eta) V(\mathrm{d}p) \to \int_M \langle \nabla f, \nabla f \rangle \,\mathrm{d}V.$$

The approach and choosing charts

Since V is a finite measure, uniform convergence of functions implies L^1 convergence. Because

of this it suffices to show that $N^2 \int_{T_pM} (f(p(1/N,\eta)) - f(p))^2 \mu_p(\mathrm{d}\eta)$ converges uniformly on M to $\langle \nabla f, \nabla f \rangle$. Now choose charts on M in exactly the same way as in section 4.1. For each $p \in M$, let (x_p, U_p) be a coordinate chart for M around p. U_p is open with respect to ρ and hence with respect to d. This means that there is some $\epsilon_p > 0$ such that $G_p := \overline{B_d(p, \epsilon_p)} \subset U_p$. Now define $O_p = B_d(p, \epsilon_p/2)$. Since M is compact, we can find $p_1, ..., p_m$ such that $M \subset \cup_i O_{p_i}$. Again, for uniform convergence it suffices to show uniform convergence on O_{p_j} for some j. Now let $j \in \{1, ..., m\}$ be fixed. Call $O := O_{p_j}, \epsilon := \epsilon_{p_j}, x := x_{p_j}, G := G_{p_j}$ and $U := U_{p_j}$. Now it suffices to show that uniformly on O:

$$N^2 \int_{M_p} (f(p(1/N,\eta)) - f(p))^2 \mu_p(\mathrm{d}\eta) \to \langle \nabla f, \nabla f \rangle (p) = \sum_{ij} \frac{\partial f}{\partial x^i}(p) \frac{\partial f}{\partial x^j}(p) g^{ij}(p).$$

We know, as in the previous section, that there exists a K such that $||\eta|| < K$ for all $\eta \in \text{supp}\mu_p$. This implies (lemma 4.2) that there is some $N_{\epsilon} \in \mathbb{N}$ such that $p(s,\eta) \in G$ for $|s| \leq 1/N$, $N \geq N_{\epsilon}$ and $||\eta|| \leq K$. Fix $N \geq N_{\epsilon}$.

Taylor expansion

As in the previous section, we proceed by using a Taylor expansion:

$$f(p(1/N,\eta)) = f(p) + \frac{1}{N} \frac{\mathrm{d}(f \circ p^{\eta})}{\mathrm{d}t}(0) + \frac{1}{2N^2} \frac{\mathrm{d}^2(f \circ p^{\eta})}{\mathrm{d}t^2}(t_{N,\eta,p}).$$

where $t_{N,\eta,p} \in (0, 1/N)$ is a number depending on N, η and p. This shows that

$$(f(p(1/N,\eta)) - f(p))^2 = \left(\frac{1}{N}\frac{\mathrm{d}(f \circ p^{\eta})}{\mathrm{d}t}(0) + \frac{1}{2N^2}\frac{\mathrm{d}^2(f \circ p^{\eta})}{\mathrm{d}t^2}(t_{N,\eta,p})\right)^2,$$

so we obtain

$$N^{2} \int (f(p(1/N,\eta)) - f(p))^{2} \mu_{p}(\mathrm{d}\eta) = \int \left(\frac{\mathrm{d}(f \circ p^{\eta})}{\mathrm{d}t}(0)\right)^{2} \mu_{p}(\mathrm{d}\eta) + \frac{1}{N} \int \frac{\mathrm{d}(f \circ p^{\eta})}{\mathrm{d}t}(0) \frac{\mathrm{d}^{2}(f \circ p^{\eta})}{\mathrm{d}t^{2}}(t_{N,\eta,p}) \mu_{p}(\mathrm{d}\eta) + \frac{1}{4N^{2}} \int \left(\frac{\mathrm{d}^{2}(f \circ p^{\eta})}{\mathrm{d}t^{2}}(t_{N,\eta,p})\right)^{2} \mu_{p}(\mathrm{d}\eta).$$
(17)

We will deal with these terms separately.

The first term

We will write the first term in local coordinates for $|t| \leq 1/N$. Using the calculations in the previous section, we see that it equals

$$\int \left(\frac{\mathrm{d}(f\circ p^{\eta})}{\mathrm{d}t}(0)\right)^2 \mu_p(\mathrm{d}\eta) = \int \left(\sum_{i=1}^n \frac{\partial f}{\partial x^i}(p)\eta^i\right)^2 \mu_p(\mathrm{d}\eta) = \int \sum_{i,j=1}^n \frac{\partial f}{\partial x^i}(p)\frac{\partial f}{\partial x^j}(p)\eta^i\eta^j\mu_p(\mathrm{d}\eta)$$
$$= \sum_{i,j=1}^n \frac{\partial f}{\partial x^i}(p)\frac{\partial f}{\partial x^j}(p)\int \eta^i\eta^j\mu_p(\mathrm{d}\eta) = \sum_{i,j=1}^n \frac{\partial f}{\partial x^i}(p)\frac{\partial f}{\partial x^j}(p)g^{ij}(p).$$

This is the expression that we want.

The rest terms

Now we have to show that the last two terms of (17) go to 0 uniformly on O. Using lemma 4.3, we see that for η in supp μ_p :

$$\begin{aligned} \left| \frac{\mathrm{d}(f \circ p^{\eta})}{\mathrm{d}t}(0) \right| &= \left| \sum_{i} \frac{\partial f}{\partial x^{i}}(p) \eta^{i} \right| &\leq \sum_{i} \left| \frac{\partial f}{\partial x^{i}}(p) \right| \left| \eta^{i} \right| \\ &\leq \sum_{i} \left| \frac{\partial f}{\partial x^{i}}(p) \right| \sqrt{g^{ii}(p)} ||\eta|| &\leq \sum_{i} \left| \frac{\partial f}{\partial x^{i}}(p) \right| \sqrt{g^{ii}(p)} K. \end{aligned}$$

As shown before in the previous section, this term can be bounded uniformly on O (since g^{ii} and $\frac{\partial f}{\partial x^i}$ are smooth and hence bounded on G which contains O). Moreover, using notation as in the previous section, we know:

$$\frac{\mathrm{d}^2(f \circ p^\eta)}{\mathrm{d}t^2} = f_{ij}p_1^i p_1^j + f_i p_2^i.$$

This shows that $\frac{d^2(f \circ p^{\eta})}{dt^2}$ and hence $(\frac{d^2(f \circ p^{\eta})}{dt^2})^2$ only contain terms that also occurred in the previous section. This means that we can uniformly bound the integrals and hence show the uniform convergence of the rest terms to 0 in exactly the same way as in the previous section. We will not repeat the analogous calculations here.

Conclusion

We conclude that the Dirichlet forms of the step processes converge for any smooth f to the Dirichlet form of Brownian motion if the following hold:

- $\sup_{p \in M} \sup_{\eta \in \operatorname{supp}\mu_n} ||\eta|| < \infty$
- $\int \eta^i \mu_p(\mathrm{d}\eta) = 0$ and $\int \eta^i \eta^j \mu_p(\mathrm{d}\eta) = g^{ij}(p)$ in each coordinate system around p.

As mentioned before, section 4.3 will deal with the question which measures satisfy these assumptions and provide some examples.

Remark on convergence of the processes

Since we have already shown the convergence of the generators and used Trotter-Kurtz in section 4.1, we already know that the corresponding processes converge (in distribution in the path space). Moreover, the convergence of the generators implies the convergence of the Dirichlet forms. The reason that we still showed the convergence of the (formal) Dirichlet forms, is to study the situation from this different perspective. In this way we can see directly from the computations how the expression for the Dirichlet form of Brownian motion emerges from a simple Taylor expansion.

In general, if there is convergence of the Dirichlet forms in a stronger sense, it is possible to directly deduce convergence in some sense of the corresponding processes. For instance there is the concept of Mosco convergence. It was introduced by Mosco. He proves in [15, Thm 2.4.1, Cor 2.6.1] that Mosco convergence of the Dirichlet forms implies strong convergence of the corresponding semigroups. This, in turn, implies weak convergence of the finite dimensional distributions of the corresponding processes. We believe that with slightly more effort Mosco convergence can be shown to hold in our situation. In some cases this might be a more suitable approach, for instance when the convergence of the generator is hard to show. For this reason it was good to explore the Dirichlet form convergence in this case, although we did not need it here.

4.3 Stepping distribution

Constraints for a stepping distribution

The question now is which distributions μ_p on T_pM satisfy assumptions 1, 2 and 3 of section 4.1. From here on we fix $p \in M$ and simply write μ for μ_p . Being compactly supported is a rather natural constraint, but the other assumptions are harder, especially since they involve local coordinates. In this section we address the question which distributions satisfy assumptions 2 and 3, i.e. for every coordinate system around p:

$$\int \eta^{i} \mu(\mathrm{d}\eta) = 0 \qquad \forall i = 1, .., n$$

$$\int \eta^{i} \eta^{j} \mu(\mathrm{d}\eta) = g^{ij} \qquad \forall i, j = 1, .., n.$$
(18)

To generalize this a bit, suppose μ satisfies the following for some c > 0 for every coordinate system:

$$\int \eta^{i} \mu(\mathrm{d}\eta) = 0 \qquad \forall i = 1, .., n$$

$$\int \eta^{i} \eta^{j} \mu(\mathrm{d}\eta) = cg^{ij} \qquad \forall i, j = 1, .., n.$$
(19)

Following the proof in the previous section, one sees directly that in this case the generators and Dirichlet forms converge to the generator and Dirichlet form, respectively, of Brownian motion that is speeded up by a factor c. We will look into this generalized situation and at the end we will see how to determine c.

Independence of coordinate systems

The following lemma shows that it suffices to check (19) in a single coordinate system. In other words, whether (19) is satisfied or not does not depend on which coordinates one works with.

<u>Lemma 4.5</u>. If (19) holds for some c > 0 and for some coordinate system (x, U) around p, then it holds for the same c for all coordinate systems around p.

Proof. Let (x, U) be a coordinate system around p for which (19) holds with c > 0 and let (y, V) be any other coordinate system around p. It suffices to show that (19) holds with the same c for y. Denote the metric matrix with respect to x by g and the one with respect to y by \hat{g} . For any $\eta \in T_p M$ define $\eta^1, ..., \eta^n$ as the coefficients of η with respect to x, so such that $\eta = \sum_i \eta^i \frac{\partial}{\partial x^i}$. Analogously let $\hat{\eta}^1, ..., \hat{\eta}^n$ be such that $\eta = \sum_i \hat{\eta}^i \frac{\partial}{\partial y^i}$. Let $J = \frac{\partial(x^1, ..., x^n)}{\partial(y^1, ..., y^n)}$. If $\eta \in T_p M$, then

$$\hat{\eta}^j = \eta(y^i) = \sum_i \eta^i \frac{\partial}{\partial x^i} y^i = \sum_i \eta^i \frac{\partial y^j}{\partial x^i}.$$

This shows that for any j

$$\int \hat{\eta}^{j} \mu(\mathrm{d}\eta) = \int \sum_{i=1}^{n} \eta^{i} \frac{\partial y^{j}}{\partial x^{i}} \mu(\mathrm{d}\eta) = \sum_{i=1}^{n} \frac{\partial y^{j}}{\partial x^{i}} \int \eta^{i} \mu(\mathrm{d}\eta) = 0,$$

since for any $i: \int \eta^i \mu(\mathrm{d}\eta) = 0$. Moreover, for any $i, j: \int \eta^i \eta^j \mu(\mathrm{d}\eta) = cg^{ij}$, so for any i, j:

$$\begin{split} \int \hat{\eta}^{i} \hat{\eta}^{j} \mu(\mathrm{d}\eta) &= \int \sum_{k=1}^{n} \eta^{k} \frac{\partial y^{i}}{\partial x^{k}} \sum_{l=1}^{n} \eta^{l} \frac{\partial y^{j}}{\partial x^{l}} \mu(\mathrm{d}\eta) = \sum_{k,l=1}^{n} \frac{\partial y^{i}}{\partial x^{k}} \frac{\partial y^{j}}{\partial x^{l}} \int \eta^{k} \eta^{l} \mu(\mathrm{d}\eta) \\ &= \sum_{k,l=1}^{n} \frac{\partial y^{i}}{\partial x^{k}} \frac{\partial y^{j}}{\partial x^{l}} cg^{kl} = c(J^{-1}G^{-1}(J^{-1})^{T})_{ij}. \end{split}$$

Since $J^{-1}G^{-1}(J^{-1})^T = J^{-1}G^{-1}(J^T)^{-1} = (J^TGJ)^{-1} = \hat{G}^{-1}$, we see that $\int \hat{\eta}^i \hat{\eta}^j \mu(d\eta) = c\hat{g}^{ij}$. We conclude that (19) holds for y with the same c.

Othogonal transformations and canonical measures

We now introduce a class of measures.

<u>Definition 4.6</u>. Let V be an inner product space and let T be a linear map $V \to V$. We call T an *orthogonal transformation* if for any $u, v \in V$: $\langle Tu, Tv \rangle = \langle u, v \rangle$.

We call a measure μ on T_pM canonical if for any orthogonal transformation T on T_pM and for any coordinate system:

$$\int \eta^{i} \mu(\mathrm{d}\eta) = \int (T\eta)^{i} \mu(\mathrm{d}\eta) \text{ and } \int \eta^{i} \eta^{j} \mu(\mathrm{d}\eta) = \int (T\eta)^{i} (T\eta)^{j} \mu(\mathrm{d}\eta).$$

<u>Remark 4.7</u>. In the same way as above, one can show that μ has the property above with respect to some coordinate system if and only if it has the property with respect to every coordinate system. Moreover, since -I always satisfies $(-I)^T G(-I) = G$, we see that $\int \eta^i \mu(\mathrm{d}\eta) = \int (-\eta)^i \mu(\mathrm{d}\eta) = \int -\eta^i \mu(\mathrm{d}\eta) = -\int \eta^i \mu(\mathrm{d}\eta)$, so $\int \eta^i \mu(\mathrm{d}\eta)$ is 0 for any canonical μ .

In words, μ is canonical if orthogonal transformations do not change the mean vector and the covariance matrix of the random variable that has distribution μ . Remark 4.7 shows that in fact the mean vector must be 0. Note that in particular measures that are invariant under orthogonal transformations are canonical, since then $\int (T\eta)^i \mu(d\eta) = \int \eta^i (\mu \circ T^{-1})(d\eta) = \int \eta^i \mu(d\eta)$ and the other equation follows analogously. However a simple example shows that the converse is not true. Let $M = \mathbb{R}$ and let μ be any non-symmetric distribution on $T_pM = \mathbb{R}$ with mean 0. The only orthogonal transformation (apart from the identity) is $t \mapsto -t$. Under this transformation the mean (which is 0) and the second moment are obviously left invariant, but μ is not symmetric, so it is not invariant. We will give an example for \mathbb{R}^n later.

If (x, U) is some coordinate system around p and $G = (g_{ij})$ is the matrix of the metric in p with respect to x, we can write a linear transformation $T : T_p M \to T_p M$ as a matrix (which we will also call T) with respect to the base $\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}$. We see that

$$\langle T\eta, T\xi \rangle = \sum_{i,j} g_{ij} (T\eta)^i (T\xi)^j = \sum_{i,j} g_{ij} \sum_k T_{ik} \eta^k \sum_l T_{jl} \xi^l = \sum_{k,l} \left(\sum_{i,j} g_{ij} T_{ik} T_{jl} \right) \eta^k \xi^l.$$

If T is orthogonal, this must equal

$$\langle \eta, v \rangle = \sum_{k,l} g_{kl} \eta^k \xi^l,$$

so we see that $g_{kl} = \sum_{i,j} g_{ij} T_{ik} T_{jl} = (T^T G T)_{kl}$ and hence $G = T^T G T$. Now for a measure μ on $T_p M$ and a coordinate system (x, U), define the vector A_{μ} and the matrix B_{μ} by $A^i_{\mu} = \int \eta^i \mu(\mathrm{d}\eta)$ and $B^{ij}_{\mu} = \int \eta^i \eta^j \mu(\mathrm{d}\eta)$. Then it is easy to show the following lemma.

<u>Lemma 4.8</u>. Let μ be a measure on T_pM . Then the following are equivalent.

- (i) μ is canonical.
- (ii) For every linear transformation T and every coordinate system (x, U): if $G = T^T G T$, then $A_{\mu} = T A_{\mu}$ and $B_{\mu} = T B_{\mu} T^T$.

Proof. (i) \Leftrightarrow (ii) because (ii) is just the definition of being canonical written in local coordinates. Indeed, we already saw that orthogonality or T translates in local coordinates to $G = T^T G T$, the other expressions follow in a similar way from the following equations:

$$A^{i}_{\mu} = \int (T\eta)^{i} \mu(\mathrm{d}\eta) = \int \sum_{k} T_{ik} \eta^{k} \mu(\mathrm{d}\eta) = \sum_{k} T_{ik} \int \eta^{k} \mu(\mathrm{d}\eta) = \sum_{k} T_{ik} A^{k}_{\mu}$$
$$B^{ij}_{\mu} = \int (T\eta)^{i} (T\eta)^{j} \mu(\mathrm{d}\eta) = \int \sum_{k} T_{ik} \eta^{k} \sum_{l} T_{jl} \eta^{l} \mu(\mathrm{d}\eta) = \sum_{k,l} T_{ik} T_{jl} \int \eta^{k} \eta^{l} \mu(\mathrm{d}\eta) = \sum_{k,l} T_{ik} T_{jl} B^{kl}_{\mu}$$

Canonical measures are stepping distributions

Now we have the following result.

Proposition 4.9. Let μ be a probability measure on T_pM . Then μ is canonical if and only if it satisfies (19) for some c > 0.

Proof. First assume that μ is canonical and let (x, U) be normal coordinates centered at p. Because of lemma 4.5 it suffices to verify (19) for x, so we need to show that $A_{\mu} = 0$ and $B_{\mu} = cG^{-1} = cI$ for some c > 0.

The fact that $A_{\mu} = 0$ is just remark 4.7. Now note that since B_{μ} is symmetric, it can be diagonalized as $TB_{\mu}T^{-1}$ where T is an orthogonal matrix (in the usual sense). This means that $T^{T} = T^{-1}$ and that $T^{T}GT = T^{T}IT = T^{T}T = I = G$, so lemma 4.8 tells us that the diagonalization equals $TB_{\mu}T^{T} = B_{\mu}$. This implies that B_{μ} is a diagonal matrix. Now for $i \neq j$ let \bar{I}^{ij} be the $n \times n$ -identity matrix with the i^{th} and j^{th} column exchanged. It is easy to see that $(\bar{I}^{ij})^T \bar{I}^{ij} = I$, so we must also have $B_{\mu} = \bar{I}^{ij}B_{\mu}(\bar{I}^{ij})^T$. The latter is B_{μ} with the i^{th} and j^{th} diagonal element exchanged. This shows that these elements must be equal. Hence all diagonal elements are equal and $B_{\mu} = cI$ for some $c \in \mathbb{R}$. Since $c = B_{\mu}^{11} = \int \eta^{1}\eta^{1}\mu(d\eta) \ge 0$, we know that $c \ge 0$. If c = 0, then $B_{\mu} = 0$, so $\mu = 0$, which is not possible. We conclude that c > 0. Conversely let (x, U) be a coordinate system with corresponding metric matrix G and assume that μ satisfies (19) for some c > 0. Let T be such that $G = T^T G T$. Then $A_{\mu} = 0 = T0 = T A_{\mu}$. We also see: $T^T G T = G \iff G = (T^T)^{-1} G T^{-1} \iff G^{-1} = T G^{-1} T^T \iff cG^{-1} = T (cG^{-1})T^T \implies B_{\mu} = T B_{\mu}T^T$ (since $B_{\mu} = cG^{-1}$), so by lemma 4.8 μ is canonical.

Now we know that if the stepping distribution is canonical (and compactly supported, uniformly on M), the generators and Dirichlet forms converge to the generator and Dirichlet form, respectively, of Brownian motion that is speeded up by some factor c > 0 (depending on μ). The question remains what this c is. The following lemma answers this question.

<u>Lemma 4.10</u>. Suppose μ satisfies (19) for some c > 0. Then $c = \frac{\int ||\eta||^2 \mu(\mathrm{d}\eta)}{n}$.

Proof. We calculate the following (with respect to some coordinate system (x, U)):

$$\begin{split} \int ||\eta||^2 \mu(\mathrm{d}\eta) &= \int \langle \eta, \eta \rangle \, \mu(\mathrm{d}\eta) = \int \left\langle \sum_i \eta^i \frac{\partial}{\partial x^i}, \sum_j \eta^j \frac{\partial}{\partial x^j} \right\rangle \mu(\mathrm{d}\eta) \\ &= \sum_{i,j} \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle \int \eta^i \eta^j \mu(\mathrm{d}\eta) = \sum_{i,j} g_{ij} c g^{ij} = c \sum_i \sum_j g_{ij} g^{ji} = c \sum_i 1 = cn. \end{split}$$

Hence $c = \frac{\int ||\eta||^2 \mu(\mathrm{d}\eta)}{n}$.

The nice part of this lemma is that the expression for c does not involve a coordinate system, only the norm (and hence inner product) of $T_p M$. In particular we see that c = 1 is equivalent to $\int ||\eta||^2 \mu(\mathrm{d}\eta) = n$. We summarize our findings in the following result.

Proposition 4.11 . A probability measure μ on T_pM satisfies (19) for a certain c > 0 if and only if it is canonical and $c = \frac{\int ||\eta||^2 \mu(\mathrm{d}\eta)}{n}$. In particular, it satisfies (18) (and, if compactly supported, hence qualifies as a stepping distribution) if and only if it is canonical and $\int ||\eta||^2 \mu(\mathrm{d}\eta) = n$.

Examples

Now let us look at some examples. By lemma 4.5, it suffices to choose a coordinate system and construct a distribution that satisfies (18) for that coordinate system. Let (x, U) be any coordinate system around some point in M with corresponding metric matrix G in that point. Let X be any random variable in \mathbb{R}^n that has mean vector 0 and covariance matrix G^{-1} (for instance let $X \sim N(0, G^{-1})$). Now let μ be the distribution of $\sum_i X^i \frac{\partial}{\partial x^i}$. Then by construction $\int \eta^i \mu(\mathrm{d}\eta) = \mathbb{E}X^i = 0$ and $\int \eta^i \eta^j \mu(\mathrm{d}\eta) = \mathbb{E}X^i X^j = \mathbb{E}X^i X^j - \mathbb{E}X^i \mathbb{E}X^j = g^{ij}$. This means that the contructed measure satisfies assumptions 2 and 3. However, it is not compactly supported, so assumption 1 does not hold.

Note that we did not need any of the theory involving canonical measures that we developed above, so let us consider an example that does need that. Let μ be the uniform distribution

on $\sqrt{n}S_pM$ (the vectors with norm \sqrt{n}). By definition of such a distribution, it is invariant under orthogonal transformations (rotations and reflections), so it is a canonical distribution. Moreover, it is clearly compactly supported (and if μ_p equals this measure for any p, they are compactly supported uniformly on M). Since also $\int ||\eta||^2 \mu(d\eta) = \int \sqrt{n^2} \mu(d\eta) = n$, we conclude that the uniform distribution on $\sqrt{n}S_pM$ can be used as a stepping distribution.

Let us conclude by showing for \mathbb{R}^n that the class of canonical probability distributions is strictly larger than the class of probability distributions that are invariant under orthogonal transformations, even with the restriction that $\int ||\eta||^2 \mu(d\eta) = n$. It suffices to find a distribution μ with mean 0 and covariance matrix I (since then μ satisfies (18) and 4.11 then tells us that μ is canonical and has $\int ||\eta||^2 \mu(d\eta) = n$) and an orthogonal T such that $\mu \neq \mu \circ T^{-1}$. Let ν be the distribution on \mathbb{R} given by $\nu = \frac{1}{5}\delta_{-2} + \frac{4}{5}\delta_{1/2}$. Then, using the natural coordinate system, $\int t\nu(dt) = \frac{1}{5}(-2) + \frac{4}{5}\frac{1}{2} = 0$ and $\int t^2\mu(dt) = \frac{1}{5}(-2)^2 + \frac{4}{5}(\frac{1}{2})^2 = 1$. Now let $\mu = \nu \times ... \times \nu$ (n times). Then we directly see that the mean vector is 0 and the covariance matrix is I. However T = -Iis an orthogonal transformation and $\mu \circ (-I)^{-1}$ equals the product of n times $\frac{1}{5}\delta_2 + \frac{4}{5}\delta_{-1/2}$, so obviously $\mu \neq \mu \circ (-I)^{-1}$.

5 Random grid

In the next section we want to study interacting particles systems on a compact and connected Riemannian manifold M. We will, more specifically, look at a process where particles perform random walks with the restriction that jumps to already occupied sites are cancelled. Our aim will be to prove that in the limit when we scale time and space appropriately, the particle densities behave according to the heat equation. In order to do all of this we need to discretize the manifold or, in other words, find a natural grid on the manifold. This turns out to be a challenging problem, because there is no obvious sequence of grids that converges to the manifold itself. The goal of this section is to find such a grid. What we want in particular is that random walks on the grids converge to Brownian motion on the manifold. We will propose a grid and show that we have convergence of the Dirichlet forms and something close to convergence of the generators. We start out with a random grid, but from this it can be deduced that it works for a lot of deterministic grids as well.

5.1 Model and motivation

Motivation

In some areas of statistics the following is known and used (see for instance [19]). Suppose we have a manifold M that is imbedded in \mathbb{R}^d for some d and we would like to recover the manifold from some observations of it, say an i.i.d. sample of uniform random elements of M. To do this we can describe the observations as a graph with as weight on the edge between any two points some semi positive kernel with bandwidth ϵ applied to the Euclidean distance between those points. Then it can be shown that the graph Laplacian of the graph that is obtained in this way converges to the Laplace-Beltrami operator on M as the number of observations goes to infinity and ϵ goes to 0. This suggests that we could define random walks on random graphs and that the corresponding generators converge to the generator of Brownian motion.

A main point of concern is the following: we prefer to view the manifold M on its own instead of imbedded in a Euclidean space. This means that we would like to use the distance that is induced by the Riemannian metric instead of the Euclidean distance. The latter is more suitable to purposes in statistics, since there the Riemannian metric on M is not known beforehand. Also, a lot is known about the behaviour of the Euclidean distance in this type of situation and not so much about the distance on the manifold. We will try to make things work in M itself.

Model

Let M be a compact and connected Riemannian manifold and let k be some semi positive kernel $[0, \infty) \to [0, \infty)$, i.e. k is non-negative and decreasing. Let $(P_i)_{i=1}^{\infty}$ be a sequence of iid uniform random elements of M (the uniform (probability) measure on M exists, because M is compact). Define

$$W_{ij}^{\epsilon} = k(d(P_i, P_j)/\epsilon)$$

as the conductance from P_i to P_j . Here *d* is the metric on *M* that is induced by the Riemannian metric (so infimum of path lengths). We can use these conductances to define a random walk on $V_N = \{P_1, .., P_N\}$ (as we did in section 2.2). Our goal now is to show that as the number of points *N* goes to infinity and the bandwidth ϵ goes to 0, the corresponding generators and Dirichlet forms converge in some sense to the generator and Dirichlet form of Brownian motion on *M*, respectively. As was mentioned before, there are already a lot of results in the case that *M* is viewed as a submanifold of \mathbb{R}^d , but these results all use the Euclidean distance. We will try to obtain similar results using the Riemannian distance.

5.2 Convergence of the Dirichlet forms

The Dirichlet form

Since we defined the random walk on a finite graph $V_N = \{P_1, ..., P_N\}$, using the 'conductances' $(W_{ij}^{\epsilon})_{i,j=1}^n$ (as in section 2.2), we can directly write down the corresponding Dirichlet form:

$$\mathcal{E}^{N}(f,g) = \sum_{i,j=1}^{N} W_{ij}^{\epsilon}(f(P_{j}) - f(P_{i}))(g(P_{j}) - g(P_{i})).$$

Note that we leave out the factor $\frac{1}{2}$ here and in the Dirichlet form of Brownian motion, so we do not have to write it every time. It clearly does not matter for the calculations. Here f, gare functions $V_N \to \mathbb{R}$. The domain of the corresponding generator is the set of all functions $V_N \to \mathbb{R}$ (as mentioned in section 2.1), so the domain of the Dirichlet form contains all these functions. This means that we can apply \mathcal{E}^N to any $f, g: V_N \to \mathbb{R}$. Whenever we apply such a Dirichlet form to functions $f, g: M \to \mathbb{R}$, we (implicitly) mean that we apply it to $f|_{V_N}, g|_{V_N}$. Using a polarization argument we can restrict to looking at the situation f = g. Since also $f(P_i) = f(P_j)$ for i = j (and recalling the notation $\mathcal{E}^N(f) = \mathcal{E}^N(f, f)$), we might as well write

$$\mathcal{E}^N(f) = \sum_{i \neq j} W_{ij}^{\epsilon} (f(P_j) - f(P_i))^2.$$

In [3] this same quantity is studied as a so called regularizer, but here the Euclidean distance is used when defining (W_{ij}^{ϵ}) . The authors of [3] also assume that the observations are sampled from some density p on \mathbb{R}^d . It is shown in this case that (when rescaled) it converges in probability to a quantity depending on the bandwidth ϵ as $N \to \infty$. The resulting expression then converges (also in a rescaled way) to $\int_{\mathbb{R}^d} ||\nabla f(x)||^2 p(x) dx$ as $\epsilon \downarrow 0$. We know the latter expression as the Dirichlet form of Brownian motion, weighted by some density p. Our aim is to show the analogous result by following the proof in [3] without using the imbedding in \mathbb{R}^d and the Euclidean norm. We will take $p = \frac{1}{V(M)}$ to be the density of the uniform distribution on M and we will not use p for this anymore to avoid confusing it with points of M.

5.2.1 The first limit: a.s. convergence as $N \to \infty$

Estimating deviation from the mean

Let f be a fixed smooth function $f: M \to \mathbb{R}$. First of all note that \mathcal{E}^N is a sum of N(N-1)random variables of the form $W_{ij}^{\epsilon}(f(P_j) - f(P_i))^2$. They are very similar, but they are obviously dependent (besides the fact that each random variable occurs twice, since $W_{ij}^{\epsilon}(f(P_j) - f(P_i))^2 =$ $W_{ji}^{\epsilon}(f(P_i) - f(P_j))^2$). However, we can follow [3] to deal with this and show convergence to the expectation of a single term. We would like to apply a result from [8] to a rescaled version of the Dirichlet form. If we define

$$U_f^{N,\epsilon} = \frac{1}{N(N-1)} \mathcal{E}^N(f) = \frac{1}{N(N-1)} \sum_{i \neq j} W_{ij}^{\epsilon} (f(P_j) - f(P_i))^2$$

it is easy to see that $U_f^{N,\epsilon}$ satisfies the description of example 5a of [8, page 24-25] with n = N, r = 2, $g(x, y) = k(d(x, y)/\epsilon)(f(y) - f(x))^2$ and $(X_1, ..., X_N) = (P_1, ..., P_N)$. Note that in the example, it is assumed that $X_k \in \mathbb{R}^d$, which is not the case in our situation. However, the only thing that is used, is that $g(X_i, X_j)$ is a real-valued random variable. This implies that it is no problem to take $X_1, ..., X_n$ from M and $g: M \times M \to \mathbb{R}$. Following the example, if we can find a and b such that $a \leq g(x, y) \leq b$ for all x, y we see that $U_f^{N,\epsilon}$ satisfies equation (5.7):

$$\mathbb{P}(U_f^{N,\epsilon} - \mathbb{E}U_f^{N,\epsilon} \geq t) \leq \mathrm{e}^{-2mt^2/(b-a)^2},$$

where m is the largest integer that does not exceed n/r. f is smooth and hence bounded on M (since M is compact). k is also bounded, because it is positive and decreasing on $[0, \infty)$. This

means that we can set $b = ||k||_{\infty} 4||f||_{\infty}^2$. Since g is non-negative, we can set a = 0. Finally m equals the largest integer not exceeding N/2. Since this is either N/2 or (N-1)/2, we can just pick the smaller one (the one that gives the roughest bound). Inserting these expressions shows that for any $t \ge 0$

$$\mathbb{P}(U_f^{N,\epsilon} - \mathbb{E}U_f^{N,\epsilon} \ge t) \le e^{-(N-1)t^2/(4||k||_{\infty}||f||_{\infty}^2)^2}.$$

By applying the same reasoning to $-U_f^{N,\epsilon}$, we obtain

$$\mathbb{P}(|U_f^{N,\epsilon} - \mathbb{E}U_f^{N,\epsilon}| \ge t) \le 2e^{-(N-1)t^2/(4||k||_{\infty}||f||_{\infty}^2)^2}.$$

Almost sure convergence

To use the estimation, we need the following lemma.

<u>Lemma 5.1</u>. Let $(Z_n)_{n=1}^{\infty}$ be a sequence of random variables. Suppose there is some c > 0 such that for every $t \ge 0$: $P(|Z_n| > t) \le e^{-cnt^2}$. Then $Z_n \to 0$ almost surely.

Proof. The definition of convergence implies that Z_n does not converge to 0 if and only if there is some $k \in \mathbb{N}$ such that $|Z_n| > \frac{1}{k}$ infinitely often. Writing i.o. for infinitely often, we see

$$\{Z_n \not\to 0\} = \{\exists k \in \mathbb{N} \text{ such that } |Z_n| \ge \frac{1}{k} \text{ i.o.}\} = \bigcup_{k \in \mathbb{N}} \{|Z_n| \ge \frac{1}{k} \text{ i.o.}\}$$

Since $\sum_{n=1}^{\infty} \mathbb{P}(|Z_n| \ge \frac{1}{k}) \le \sum_{n=1}^{\infty} e^{-cn(1/k)^2} < \infty$, the Borel-Cantelli lemma says that

$$\mathbb{P}(|Z_n| \ge \frac{1}{k} \text{ i.o.}) = 0.$$

Now we obtain:

$$\mathbb{P}(Z_n \not\to 0) = \mathbb{P}(\bigcup_{k \in \mathbb{N}} \{ |Z_n| \ge \frac{1}{k} \text{ i.o.} \}) \le \sum_{k \in \mathbb{N}} \mathbb{P}(|Z_n| \ge \frac{1}{k} \text{ i.o.}) = 0.$$

This means that $\mathbb{P}(Z_n \to 0) = 1$, so $Z_n \to 0$ almost surely.

Applying this lemma to our situation directly shows that $U_f^{N,\epsilon} - \mathbb{E}U_f^{N,\epsilon} \to 0$ almost surely as $N \to \infty$. Now, using the fact that for any pair (i, j) with $i \neq j$ the distribution of $W_{ij}^{\epsilon}(f(P_j) - f(P_i))^2$ is the same, we see that for any N

$$\mathbb{E}U_{f}^{N,\epsilon} = \frac{1}{N(N-1)} \sum_{i \neq j} \mathbb{E}W_{ij}^{\epsilon} (f(P_{j}) - f(P_{i}))^{2} = \mathbb{E}W_{ij}^{\epsilon} (f(P_{j}) - f(P_{i}))^{2},$$

where i, j are arbitrary such that $i \neq j$. This means that $\mathbb{E}U_f^{N,\epsilon}$ is constant in N, which (together with the a.s. convergence result above) implies that $U_f^{N,\epsilon} \to \mathbb{E}W_{ij}^{\epsilon}(f(P_j) - f(P_i))^2$ almost surely. The expectation equals

$$a_f^{\epsilon} := \mathbb{E}W_{ij}^{\epsilon}(f(P_j) - f(P_i))^2 = \int_M \int_M k(d(p,q)/\epsilon)(f(q) - f(p))^2 \bar{V}(\mathrm{d}q)\bar{V}(\mathrm{d}p)$$
(20)

where \bar{V} denotes the uniform measure $\frac{V}{V(M)}$ on M.

Interpretation of the limit

Note that (20) is the (formal) Dirichlet form of a random walk on M that jumps from p to a point on M according to the density $k(d(p, \cdot))$. $k(d(p, \cdot))$ is generally not a probability density, but one can interpret it as a probability density times a scaling factor. This factor is its integral, which exists and is positive since k is non-negative and bounded and M is compact. The

process jumps according to the probability density with the scaling factor as its jumping rate. Of course the process must be reversible for this to be a true Dirichlet form.

The convergence of $\frac{1}{N(N-1)}\mathcal{E}^N(f)$ also implies the convergence of $\frac{1}{N^2}\mathcal{E}^N(f)$. Indeed,

$$\frac{1}{N^2}\mathcal{E}^N(f) = \left(1 - \frac{1}{N}\right) \frac{1}{N(N-1)} \mathcal{E}^N(f) \to \int_M \int_M k(d(p,q)/\epsilon) (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \bar{V}(\mathrm{d}p) \text{ a.s.}$$

5.2.2 The second limit: $\epsilon \downarrow 0$

From here we cannot follow the method of [3] anymore, since their calculations are in \mathbb{R}^d . We would like to show that

$$\epsilon^{-2-n} \int_M \int_M k(d(p,q)/\epsilon) (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \bar{V}(\mathrm{d}p) \to D^k \int_M \langle \nabla f, \nabla f \rangle \mathrm{d}V$$
(21)

as $\epsilon \downarrow 0$, where D^k is a constant that depends on the kernel (and, of course, on M). Before we do this, we restrict ourselves to a specific situation.

Choice of kernel k

At this point we introduce a specific kernel, namely $k = \mathbb{1}_{[0,1]}$ (we will show that this can be generalized later). We denote for $p \in M, r > 0$: $B_d(p,r) = \{q \in M : d(p,q) \leq r\}$. Then we can write

$$\int_{M} k(d(u,v)/\epsilon) (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) = \int_{B_d(p,\epsilon)} (f(q) - f(p))^2 \bar{V}(\mathrm{d}q).$$
(22)

For this choice of k, we can say more about the random walk that corresponds to the resulting Dirichlet form. It jumps (with a certain rate) from p to a point that is uniformly chosen from the open ball around p with radius ϵ (with respect to d).

Interchanging the limit and the first integral

Rewriting the left hand side of (21), using (22) and writing $D = D^{\mathbb{1}_{[0,1]}}$ shows that we want:

$$\epsilon^{-2-n} \int_M \int_{B_M(p,\epsilon)} (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \frac{1}{V(M)} V(\mathrm{d}p) \to D \int_M < \nabla f, \nabla f > \mathrm{d}V.$$

Note that rescaling by V(M) was necessary because the measure V that is used in the Dirichlet form of Brownian motion is not a probability measure. Now we can multiply by V(M) and we call C = V(M)D. If we can interchange the first integral and the limit, it suffices to show that

$$\epsilon^{-2-n} \int_{B_d(p,\epsilon)} (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \to C < \nabla f, \nabla f > (p) \text{ as } \epsilon \downarrow 0.$$
(23)

This means that we first need to show that

$$\lim_{\epsilon \downarrow 0} \epsilon^{-2-n} \int_M \int_{B_d(p,\epsilon)} (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \frac{1}{V(M)} V(\mathrm{d}p)$$

=
$$\int_M \lim_{\epsilon \downarrow 0} \epsilon^{-2-n} \int_{B_d(p,\epsilon)} (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \frac{1}{V(M)} V(\mathrm{d}p).$$

We will show that $|\epsilon^{-2-n} \int_{B_d(p,\epsilon)} (f(q) - f(p))^2 \overline{V}(dq)|$ is bounded by some constant c for ϵ small enough. Since we are dealing with the finite measure V, the integral of the constant function c/V(M) is finite, so then we have bounded the integrand by an integrable function. The rest of this section will be devoted to showing that the integrands actually converge (and to what). The dominated convergence theorem then tells us that the integral and the limit can be exchanged. We start with the following lemmas.

<u>Lemma 5.2</u>. Let M be a smooth and compact Riemannian manifold and let f be a smooth function $M \to \mathbb{R}$. Then f is Lipschitz continuous.

Proof. Let M and f be fixed as above. Since M is compact and any locally Lipschitz function on a compact space is Lipschitz, it suffices to show that f is locally Lipschitz. Let $p \in M$. Let (x, U) be a chart around p. Let $G = \overline{B_d(p, \epsilon)}$ and $O = B_d(p, \epsilon/4)$ with ϵ small enough such that $G \subset U$. Moreover, by prop 3.15, ϵ can be chosen small enough such that any two points in Gare joined by a unique geodesic with length smaller than some number $\delta > 0$. We claim that fis Lipschitz on O.

Pick any points $q, r \in O$. Then there is a unique geodesic γ from q to r of length $< \delta$. Since M is compact, by theorem 3.13 (and the remark after it) it is geodesically complete. This means that q and r can be joined by a geodesic of minimal length. Since γ is the only geodesic between q and r of length $< \delta$, it must be the unique minimal geodesic. This means that the length of the geodesic equals d(q, r). Let $\eta \in T_q M$ be the unit tangent vector at q in the direction of γ . Then $q(\cdot, \eta)$ is the minimal geodesic from q to r with unit speed (it is γ rescaled to unit speed). This means that $r = q(d(q, r), \eta)$ (the geodesic arrives at r after time d(q, r) since it travels at unit speed and the distance is d(q, r)). Now we can use a Taylor expansion:

$$f(r) = f(q(d(q, r), \eta)) = f(q) + d(q, r) \frac{\mathrm{d}f \circ q^{\eta}}{\mathrm{d}t}(s)$$

with $s \in [0, d(p, q)]$. This shows that

$$|f(r) - f(q)| = d(q, r) \left| \frac{\mathrm{d}f \circ q^{\eta}}{\mathrm{d}t}(s) \right|.$$

Since the length of the geodesic from q to r is $d(q,r) \leq \epsilon/2$ (recall that $q, r \in O$), the geodesic cannot leave G between q and r. This means that it can be shown with exactly the same methods as in section 4.1 that $\left|\frac{\mathrm{d}f \circ q^{\eta}}{\mathrm{d}t}(s)\right|$ is bounded for $q \in O, s \in [0, \epsilon/2]$ and η such that $||\eta|| = 1$ by some number L. We conclude that for all $q, r \in O$:

$$|f(r) - f(q)| \le Ld(q, r).$$

This proves that f is locally Lipschitz and hence Lipschitz on M.

<u>Lemma 5.3</u>. Let M be a smooth and compact Riemannian manifold and let T be a 2-tensor field on M. Then T is bounded.

Proof. We want to show that there exist $a, b \in \mathbb{R}$ such that for all $v \in TM$: $a \langle v, v \rangle \leq T(v, v) \leq b \langle v, v \rangle$. Because of the multilinearity of every term, it suffices to show this for every $v \in TM$ with ||v|| = 1, so we need a, b such that for those $v: a \leq T(v, v) \leq b$.

Now we choose charts. For any $p \in M$ let (x_p, U_p) be a chart around p. Let $G_p \subset U_p$ be closed and let $O_p \subset G_p$ be open (we have seen before that this can be done). Now let $p_1, ..., p_m$ be such that $O_{p_1}, ..., O_{p_m}$ cover M (this can be done since M is compact). If we can find a_i, b_i such that for every i: $a_i \leq T(v, v) \leq b_i$ for v in $TM|_{O_{p_i}}$ with ||v|| = 1, we can set $a = \min_i a_i$ and $b = \max_i b_i$. So let $1 \leq j \leq n$ be fixed and call $O := O_{p_j}, G := G_{p_j}$ and $(x, U) := (x_{p_j}, U_{p_j})$. Let $p \in O$ and $v \in T_pM$ such that ||v|| = 1. We can write $T = T_{kl} dx^k \otimes dx^l$ and $v = v^r \partial_r$. Using lemma 4.3, we see:

$$|T(v,v)| = |T_{kl} dx^{k} \otimes dx^{l} (v^{r} \partial_{r}, v^{s} \partial_{s})| = |T_{kl} v^{k} v^{l}| = |T_{kl}| |v^{k}| |v^{l}|$$

$$\leq |T_{kl}| ||v|| \sqrt{g^{kk}} ||v|| \sqrt{g^{ll}} = |T_{kl}| \sqrt{g^{kk}} \sqrt{g^{ll}}.$$

Since T is smooth, so are it components T_{kl} . This means that they are continuous and hence bounded on G (since G is compact). g^{kk} is also smooth and hence continuous for any k. So $\sqrt{g^{kk}}$ is bounded on G for any k. In particular T_{kl} and $\sqrt{g^{kk}}$ are bounded on O, so T(v,v) is bounded for $v \in TM|_O$. This shows that we can find a_j, b_j such that $a_j \leq |T(v,v)| \leq b_j$ for all $v \in TM|_O = TM|_{O_{p_i}}$, which is what we wanted to show. Now we can prove what we would like.

<u>Lemma 5.4</u>. There is an $\epsilon' > 0$ and a C > 0 such that for all $0 < \epsilon < \epsilon'$ and for all $p \in M$:

$$\left| \epsilon^{-2-n} \int_{B_d(p,\epsilon)} (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \right| < C.$$

Proof. By lemma 5.2, f is Lipschitz. Call the corresponding Lipschitz constant K. Now we see:

$$\begin{aligned} \left| \int_{B_d(p,\epsilon)} (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \right| &\leq \int_{B_d(p,\epsilon)} (Kd(p,q))^2 \bar{V}(\mathrm{d}q) \\ &\leq \int_{B_d(p,\epsilon)} K^2 \epsilon^2 \bar{V}(\mathrm{d}q) = \frac{V(B_d(p,\epsilon))K^2 \epsilon^2}{V(M)}. \end{aligned}$$

Since M is compact and Ricci curvature is a smooth 2-tensor field, lemma 5.3 tells us that the Ricci curvature on M is bounded from below. This means that there is an L such that Ric $\geq (n-1)L$. Let M^L be the complete, n-dimensional, simply connected manifold with constant sectional curvature L.¹ Then the Bishop-Gromov inequality says that $V(B_d(p,\epsilon)) \leq$ $V(B^L(p^L,\epsilon))$ where $p \in M$ and $p^L \in M^L$ are arbitrary and B^L denotes a ball with respect to the Riemannian distance in M^L . So fix $p^L \in M^L$. Since $B^L(p^L,\epsilon)$ is a ball in an n-dimensional manifold, its volume is $V(B^L(p^L,\epsilon)) = O(\epsilon^n)$ (see, for instance, [24, Thm 2.5]). This means that there is some $C^L > 0$ and $\epsilon' > 0$ such that for $0 < \epsilon < \epsilon'$: $V(B^L(p^L,\epsilon)) < C^L\epsilon^n$. Now we see for $0 < \epsilon < \epsilon'$ and for all $p \in M$:

$$\begin{aligned} \left| \epsilon^{-2-n} \int_{B_d(p,\epsilon)} (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \right| &\leq \epsilon^{-2-n} \frac{V(B_d(p,\epsilon))K^2 \epsilon^2}{V(M)} \leq \epsilon^{-2-n} \frac{V(B^L(p^L,\epsilon))K^2 \epsilon^2}{V(M)} \\ &\leq \epsilon^{-2-n} \frac{C^L \epsilon^n K^2 \epsilon^2}{V(M)} = \frac{C^L K^2}{V(M)}. \end{aligned}$$

Integral over tangent space

Now that we know that the limit can be pulled through the first integral, what remains to be shown is (23). Denote for $\eta \in T_pM, r > 0$: $B_p(\eta, r) = \{\xi \in T_pM : ||\xi - \eta|| \le r\}$ (not to be confused with B_ρ , which is a ball in M with respect to the original metric ρ). For ϵ small enough we know that $\exp_p : T_pM \supset B_p(0, \epsilon) \to B_d(p, \epsilon) \subset M$ is a diffeomorphism. We want to use this to write the integral above as an integral over $B_p(0, \epsilon) \subset T_pM$:

$$\int_{B_d(p,\epsilon)} (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) = \int_{B_p(0,\epsilon)} (f(\exp_p(\eta)) - f(p))^2 \bar{V} \circ \exp(\mathrm{d}\eta) \\
= \int_{B_p(0,1)} (f(\exp_p(\epsilon\eta)) - f(p))^2 \bar{V} \circ \exp \circ \lambda_\epsilon(\mathrm{d}\eta). \quad (24)$$

This means we integrate with respect to the measure $\overline{V} \circ \exp \circ \lambda_{\epsilon}$, where λ_{ϵ} denotes multiplication with ϵ .

Determining the measure $\bar{V} \circ \exp \circ \lambda_{\epsilon}$

Since $B_p(0,\epsilon)$ is a star-shaped open neighbourhood of 0, we see that for ϵ small enough $V_{\epsilon} := B_d(p,\epsilon) = \exp_p(B_p(0,\epsilon))$ is a normal neighbourhood of p, so there exists a normal coordinate system (x, V_{ϵ}) that is centered at p. We interpret for $v \in \mathbb{R}^n$ $v_p \in T_pM$ as $\sum_i v_i \frac{\partial}{\partial x^i}$.

¹Space forms are complete, connected Riemannian manifolds with constant sectonial curvature. We will not go into their existence and properties here. See for instance [25] for a survey of this topic.

Consequently, when we write A_p for some subset A of \mathbb{R}^n , we mean $\{v_p : v \in A\}$. Since the basis $W = \left(\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}\right)$ is orthogonal in T_pM , it is easy to see that $\phi := v_p \mapsto v$ preserves the inner product and is an isomorphism of inner product spaces. Indeed,

$$||v_p||^2 = \langle v_p, v_p \rangle = (v_p)^i (v_p)^j g_{ij} = \sum_{ij} v^i v^j \delta^i_j = \sum_i (v^i)^2 = ||v||^2.$$

In particular $B_{\mathbb{R}^n}(0,\epsilon)_p = B_p(0,\epsilon)$ (where $B_{\mathbb{R}^n}$ denotes a ball in \mathbb{R}^n with respect to the Euclidean metric). We can use this in the following lemma, which tells us more about $\overline{V} \circ \exp \circ \lambda_{\epsilon}$.

<u>Lemma 5.5</u>. There exist $\epsilon' > 0$ and a function $h : B_{\mathbb{R}^n}(0, \epsilon') \to \mathbb{R}$ such that $h(t) = O(||t||^2)$ and for all $0 < \epsilon < \epsilon'$: $\bar{V} \circ \exp \circ \lambda_{\epsilon} = \epsilon^n \left(\frac{1+h(\epsilon t)}{V(M)} \mathrm{d}t^1 ..\mathrm{d}t^n\right) \circ \phi$ on $B_p(0, 1)$.

Proof. Let ϵ' be small enough such that the considerations above the lemma hold and let $\epsilon < \epsilon'$. For clarity of the proof, we first separately prove the following statement.

<u>Claim</u>: $x \circ \exp = \phi$ on $B_{\mathbb{R}^n}(0, \epsilon)_p$.

Proof. The geodesics through p are straight lines with respect to x, so they are of the form $x(\gamma(t)) = ta + b$ with $a, b \in \mathbb{R}^n$. For $\eta = \sum_i \eta^i \frac{\partial}{\partial x^i}$, the geodesic starting at p with tangent vector η at p should satisfy b = x(p) = 0 and $a_i = \eta^i$ for all i, so we see $\gamma^k = t\eta^k$. For $q \in B_d(p, \epsilon)$, we see $x^k(\exp(x(q)_p)) = 1 * x^k(q) = x^k(q)$, so $\exp(x(q)_p) = q$. This also shows that $x \circ \exp(v_p) = v$ for $v \in B_{\mathbb{R}^n}(0, \epsilon)$ (since x is invertible), which gives an identification

$$x \circ \exp: T_p M \supset B_{\mathbb{R}^n}(0,\epsilon)_p \to B_{\mathbb{R}^n}(0,\epsilon) \subset \mathbb{R}^n$$

which is the restriction of ϕ to $B_{\mathbb{R}^n}(0,\epsilon)_p$. This situation is sketched in figure 5.

Now we will first use the definition of integration to see what the measure is in coordinates (so it becomes a measure on a subset of \mathbb{R}^n). Then we will use the claim above: we will pull the measure on \mathbb{R}^n back to T_pM using ϕ .

On (x, V_{ϵ}) the volume measure is given by $\sqrt{\det G} dx^1 \wedge ... \wedge dx^n$. According to [24, Cor 2.3], $\sqrt{\det G}$ can be expanded (in normal coordinates) as 1 + h(x) where h is such that $h(x) = O(||x||^2)$. Now the measure can be written in local coordinates on $B_{\mathbb{R}^n}(\epsilon')$ as $(1 + h(x))dx^1 \wedge ... \wedge dx^n$, so the uniform measure is $\frac{1+h(x)}{V(M)}dx^1 \wedge ... \wedge dx^n$. This yields the measure $\bar{V} \circ x^{-1} = \frac{1+h(t)}{V(M)}dt^1...dt^n$ on $x(V_{\epsilon'}) = B_{\mathbb{R}^n}(0,\epsilon')$. We have on $B_{\mathbb{R}^n}(0,1)_p$:

$$\bar{V} \circ \exp \circ \lambda_{\epsilon} = (\bar{V} \circ x^{-1}) \circ (x \circ \exp) \circ \lambda_{\epsilon}.$$

According to the claim above, $x \circ \exp$ is a restriction of ϕ , so we can replace it by ϕ . Since this map is linear, it can be interchanged with λ_{ϵ} , which yields (inserting what we found before and since $\epsilon < \epsilon'$):

$$\left(\frac{1+h(t)}{V(M)}\mathrm{d}t^1..\mathrm{d}t^n\right)\circ\lambda_{\epsilon}\circ\phi=\left(\frac{\epsilon^n(1+h(\epsilon t))}{V(M)}\mathrm{d}t^1..\mathrm{d}t^n\right)\circ\phi.$$

In the last step we interpret $\frac{\epsilon^n(1+h(\epsilon t))}{V(M)} dt^1 ... dt^n$ as a measure on $B_{\mathbb{R}^n}(0,1)$ and this last step is then just a transformation of measures on \mathbb{R}^n . This yields the expression that we want. \Box

<u>Remark 5.6</u>. We used [24, Cor 2.3] in the proof above. In these notes the expansion of $\sqrt{\det G}$ is calculated around a point p in local coordinates centered around p. As can be seen, there are no linear terms in the expansion. The coefficients for the quadratic terms are coefficients of the Ricci curvature of M in p. This implies that the way that the uniform distribution on a ball around p in M is pulled back to the tangent space via the exponential map depends on the curvature of M in p. In particular, if there is no curvature, M is locally isomorphic to a neighbourhood in \mathbb{R}^n so the same thing happens as in \mathbb{R}^n . This means that we get a uniform distribution on a ball around 0 in the tangent space.


Figure 5: The situation in lemma 5.5. On $B_p(0,\epsilon)$: $x \circ \exp = \phi$. The uniform measure on $B_d(p,\epsilon)$ is moved via x to $B_{\mathbb{R}^n}(0,\epsilon)$ using the formula $\sqrt{\det Gt_1..t_n}$. This measure can then be pulled back to $B_p(0,\epsilon)$ using ϕ . Since ϕ is an inner product space isomorphism, it will be easy to deal with orthogonal transformations later, in lemma 5.7.

A canonical part plus a rest term

Now define

$$\mu = \left(\frac{1}{V(M)} dt^1 .. dt^n\right) \circ \phi \quad \text{and} \quad \mu_R = \left(\frac{h(\epsilon t)}{V(M)} dt^1 .. dt^n\right) \circ \phi$$

on $B_p(0,1)$ and 0 everywhere else. Then the lemma implies that (24) equals

$$\int_{B_p(0,1)} (f(\exp_p(\epsilon\eta)) - f(p))^2 \epsilon^n (\mu + \mu_R) (\mathrm{d}\eta) = \epsilon^n \int_{T_pM} (f(p(\epsilon,\eta)) - f(p))^2 (\mu + \mu_R) (\mathrm{d}\eta).$$

Recall that $p(\epsilon, \eta)$ is just notation for following the geodesic from p in the direction of η for time ϵ . In this way we transformed the integral to one that we are familiar with from section 4.2 since we wrote the stepping distribution as a distribution on the tangent space instead of a distribution on M. To use the theory that we obtained in that section, we need the following lemma.

<u>Lemma 5.7</u>. μ is canonical. Moreover $\int_{T_pM} ||\eta||^2 \mu(\mathrm{d}\eta) = \frac{2\pi^{n/2}}{V(M)\Gamma(n/2)(n+2)}$.

Proof. Define $\nu = \frac{1}{V(M)} dt^1 ... dt^n$ on $B_{\mathbb{R}^n}(0, 1)$ and 0 everywhere else. Then we can write $\mu = \nu \circ \phi$. Since ϕ preserves the inner product, the measure μ behaves the same with respect to orthogonal transformations in $T_p M$ as ν with respect to orthogonal transformations in \mathbb{R}^n . Since ν is obviously preserved under such transformations, so is μ . This shows that μ is canonical. Now we calculate the corresponding constant.

$$\begin{split} \int_{T_pM} ||\eta||^2_{T_pM} \mu(\mathrm{d}\eta) &= \int_{T_pM} ||v_p||^2_{T_pM} \mu(\mathrm{d}v_p) = \int_{\mathbb{R}^n} ||\phi^{-1}(v)||^2_{T_pM} \nu(\mathrm{d}v) = \int_{\mathbb{R}^n} ||v||^2_{\mathbb{R}^n} \nu(\mathrm{d}v) \\ &= \frac{1}{V(M)} \int_{B_{\mathbb{R}^n}(0,1)} ||v||^2_{\mathbb{R}^n} \mathrm{d}v = \frac{1}{V(M)} \int_0^1 r^2 \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1} \mathrm{d}r = \frac{1}{V(M)} \frac{2\pi^{n/2}}{\Gamma(n/2)} \frac{1}{n+2} \end{split}$$

The first step was just writing the integral with respect to the coordinates for which we defined μ . The second step holds because $\mu = \nu \circ \phi$. The third uses the fact that ϕ preserves the norm. The but-one-last step is a change of coordinates in \mathbb{R}^n using the fact that ||v|| is constant on spheres around the origin. Here $\frac{2\pi^{n/2}}{\Gamma(n/2)}r^{n-1}$ is the area of rS_{n-1} .

5.2.3 Conclusion and generalization

We collect everything and summarize our findings in the following proposition.

<u>Proposition 5.8</u>. Let $(P_i)_{i=1}^{\infty}$ be as above. Let $k = \mathbb{1}_{[0,1]}$. Then

$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \frac{1}{N^2 \epsilon^{2+n}} \mathcal{E}^N(f) = \frac{2\pi^{n/2}}{V(M)^2 \Gamma(n/2) n(n+2)} \int_M \langle \nabla f, \nabla f \rangle \, \mathrm{d}V,$$

where the limit of $N \rightarrow \infty$ is an almost sure limit.

Proof. First of all it follows from above that

$$\lim_{N \to \infty} \frac{1}{N^2 \epsilon^{2+n}} \mathcal{E}^N(f) = \frac{1}{\epsilon^{2+n}} \int_M \int_M k(d(p,q)/\epsilon) (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \bar{V}(\mathrm{d}p) \text{ a.s.}$$
(25)

By bounding the integrand of the outer integral and using dominated convergence, we saw that we can pull the limit into the integral. We also rewrote the measure to deal with the volume measure instead of the uniform measure. We thus obtain:

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon^{2+n}} \int_{M} \int_{M} k(d(p,q)/\epsilon) (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \bar{V}(\mathrm{d}p)$$

$$= \int_{M} \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon^{2+n}} \int_{M} k(d(p,q)/\epsilon) (f(q) - f(p))^2 \bar{V}(\mathrm{d}q) \frac{1}{V(M)} V(\mathrm{d}p). \tag{26}$$

We can now calculate the integrand.

$$\int_{M} k(d(p,q)/\epsilon) (f(q) - f(p))^{2} \bar{V}(\mathrm{d}q) = \epsilon^{n} \int_{T_{p}M} (f(p(\epsilon,\eta)) - f(p))^{2} (\mu + \mu_{R})(\mathrm{d}\eta)$$

= $\epsilon^{n} \int_{T_{p}M} (f(p(\epsilon,\eta)) - f(p))^{2} \mu(\mathrm{d}\eta) + \epsilon^{n} \int_{T_{p}M} (f(p(\epsilon,\eta)) - f(p))^{2} \mu_{R}(\mathrm{d}\eta).$

From the results in section 4.2 and 4.3 and lemma 5.7, we see

$$\begin{split} \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon^{2+n}} \epsilon^n \int_{T_p M} (f(p(\epsilon,\eta)) - f(p))^2 \mu(\mathrm{d}\eta) &= \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon^2} \int_{T_p M} (f(p(\epsilon,\eta)) - f(p))^2 \mu(\mathrm{d}\eta) \\ &= \frac{2\pi^{n/2}}{V(M)\Gamma(n/2)n(n+2)} < \nabla f, \nabla f > (p). \end{split}$$

Let $\epsilon'', K > 0$ such that $\epsilon'' < \epsilon'$ and $|h(s)| < K||s||^2$ for $s \in B_{\mathbb{R}^n}(0, \epsilon'')$ (where both ϵ' and h are from lemma 5.5). Note that for $\epsilon < \epsilon''$:

$$|\mu_R| \le \left(\sup_{t \in B_{\mathbb{R}^n}(0,1)} |h(\epsilon t)|\right) \mu \le \left(\sup_{t \in B_{\mathbb{R}^n}(0,1)} K||\epsilon t||^2\right) \mu = \left(\sup_{t \in B_{\mathbb{R}^n}(0,1)} K\epsilon^2 ||t||^2\right) \mu = K\epsilon^2 \mu.$$
we see for the other integral:

Now we see for the other integral:

$$\begin{split} &\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon^{2+n}} \epsilon^n \left| \int_{T_p M} (f(p(\epsilon,\eta)) - f(p))^2 \mu_R(\mathrm{d}\eta) \right| \leq \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon^2} \int_{T_p M} (f(p(\epsilon,\eta)) - f(p))^2 |\mu_R|(\mathrm{d}\eta) \\ \leq &\lim_{\epsilon \downarrow 0} K \epsilon^2 \frac{1}{\epsilon^2} \int_{T_p M} (f(p(\epsilon,\eta)) - f(p))^2 \mu(\mathrm{d}\eta) = \lim_{\epsilon \downarrow 0} K \epsilon^2 * \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon^2} \int_{T_p M} (f(p(\epsilon,\eta)) - f(p))^2 \mu(\mathrm{d}\eta) \\ = &0 * \frac{2\pi^{n/2}}{V(M)\Gamma(n/2)n(n+2)} < \nabla f, \nabla f > (p) = 0. \end{split}$$

.

Combining everything above shows that (26) equals:

$$\begin{split} &\int_{M} \left(\frac{2\pi^{n/2}}{V(M)\Gamma(n/2)n(n+2)} < \nabla f, \nabla f > (p) + 0 \right) \frac{1}{V(M)} V(\mathrm{d}p) \\ &= \frac{2\pi^{n/2}}{V(M)^{2}\Gamma(n/2)n(n+2)} \int_{M} < \nabla f, \nabla f > \mathrm{d}V. \end{split}$$

Note that we first showed that when taking the limit in N (and dividing by N^2), the Dirichlet forms (using $k = \mathbb{1}_{[0,1]}$) converge to the (formal) Dirichlet form of the random walk that jumps from each point p to a point that is chosen uniformly from $B_d(p, \epsilon)$ (rescaled by V(M)). Then we showed that doing this corresponds to choosing a vector according to some distribution from the unit ball around the origin of T_pM and following it for time ϵ . We showed that this distribution is a canonical μ plus a rest term μ_R and we calculated the corresponding time speed constant of μ . Then we just applied the theory from the previous section and showed that the rest term vanishes in the limit. We summarize part of this in the following corollary.

Corollary 5.9. In the situation above, as $N \to \infty$, the Dirichlet forms of the corresponding processes converge (almost surely) to the (formal) Dirichlet form of a random walk as defined in section 4.1. The corresponding stepping distribution is μ as defined above plus a rest term (which disappears in the limit when $\epsilon \downarrow 0$).

Generalization of the kernel k

Note that for the first limit we only needed that k is bounded. After that we chose $k = \mathbb{1}_{[0,1]}$ to deal with the second limit, but this can be generalized. First of all, note that the calculations above hold for $k = \mathbb{1}_{[0,s]}$ for any s > 0 (although the limiting process will obviously have a different speed). Now, as above, we write D^k for the constant such that

$$\epsilon^{-2-n} \int_{M} \int_{M} k(d(p,q)/\epsilon) (f(q) - f(p))^{2} \bar{V}(\mathrm{d}q) \bar{V}(\mathrm{d}p) \to D^{k} \int_{M} \langle \nabla f, \nabla f \rangle \,\mathrm{d}V.$$
(27)

Also denote

$$I^{\epsilon}(k) := \epsilon^{-2-n} \int_{M} \int_{M} k(d(p,q)/\epsilon) (f(q) - f(p))^{2} \bar{V}(\mathrm{d}q) \bar{V}(\mathrm{d}p), \quad J = \int_{M} \langle \nabla f, \nabla f \rangle \,\mathrm{d}V.$$

If f is constant, there is not much prove, so assume that f is not constant. It is easy to see that this implies that there is some point where ∇f is not 0, so $\langle \nabla f, \nabla f \rangle$ is not 0. Since ∇f is a smooth section of TM, we know $p \to \langle \nabla f, \nabla f \rangle$ is continuous, so there is an non-empty, open set where $\langle \nabla f, \nabla f \rangle$ is not 0. Since V gives positive mass to such sets (and $\langle \nabla f, \nabla f \rangle \ge 0$), we conclude $J \neq 0$. We will need this later. Now suppose we have (27) for functions $k, l : [0, \infty) \to \mathbb{R}$. Then we see for $\alpha, \beta \in \mathbb{R}$:

$$I^{\epsilon}(\alpha k + \beta l) = \alpha I^{\epsilon}(k) + \beta I^{\epsilon}(l) \to \alpha D^{k}J + \beta D^{l}J = (\alpha D^{k} + \beta D^{l})J,$$

so we have (27) for $\alpha k + \beta l$ and $D^{\alpha k+\beta l} = \alpha D^k + \beta D^l$. So linear combinations will also work (note, however, that if a function is not positive and decreasing, it does not make much sense as a kernel). For $a, b \geq 0$: $\mathbb{1}_{(a,b]}$ equals $\mathbb{1}_{[0,b]} - \mathbb{1}_{[0,a]}$, so using linearity we have the same result for $\mathbb{1}_{(a,b]}$. Since adding $\{a\}$ does not change the integrals it also works for $\mathbb{1}_{[a,b]}$. Linear combinations of these are the step functions $[0,\infty) \to \mathbb{R}$, so we can choose k to be any step function and (27) holds. Now one can try to approximate general functions by step functions. We have the following proposition.

Proposition 5.10. Let k be a function $[0, \infty) \to \mathbb{R}$. Suppose there are step functions $(k_m)_{m=1}^{\infty}$ such that $k_m \to k$ uniformly as $m \to \infty$. Suppose that there is an α such that $\operatorname{suppk}, \operatorname{suppk}_m \subset [0, \alpha]$ for all m. Also assume that $\lim_{m\to\infty} D^{k_m}$ exists. Then (27) holds for k. Moreover, $D^k = \lim_{m\to\infty} D^{k_m}$.

Proof. Let k, $(k_m)_{m=1}^{\infty}$ and α be as above. Since the supports of k and all the k_m are contained in $[0, \alpha]$, we have for all $t \ge 0$: $|k(t) - k_m(t)| \le ||k - k_m||_{\infty} \mathbb{1}_{[0,\alpha]}$. We thus see that

$$\begin{aligned} |I^{\epsilon}(k) - I^{\epsilon}(k_{m})| &\leq \epsilon^{-2-n} \int \int |k - k_{m}| (d(p,q)/\epsilon) (f(q) - f(p))^{2} \bar{V}(\mathrm{d}q) \bar{V}(\mathrm{d}p) \\ &\leq ||k - k_{m}||_{\infty} \epsilon^{-2-n} \int \int \mathbb{1}_{[0,\alpha]} (d(p,q)/\epsilon) (f(q) - f(p))^{2} \bar{V}(\mathrm{d}q) \bar{V}(\mathrm{d}p) \\ &= ||k - k_{m}||_{\infty} I^{\epsilon} (\mathbb{1}_{[0,\alpha]}). \end{aligned}$$

Since $I^{\epsilon}(\mathbb{1}_{[0,\alpha]})$ converges to $D^{\mathbb{1}_{[0,\alpha]}}J$ as ϵ goes to 0, we can find an $\epsilon' > 0$ such that for all $\epsilon < \epsilon' : |I^{\epsilon}(\mathbb{1}_{[0,\alpha]})| < |D^{\mathbb{1}_{[0,\alpha]}}J| + 1 =: K$. This means that for all $\epsilon < \epsilon'$ and for all m we have $|I^{\epsilon}(k) - I^{\epsilon}(k_m)| < ||k - k_m||_{\infty}K$.

Now let $\hat{\epsilon} > 0$. Let M_1 be such that for all $m \ge M_1$ and for all $\epsilon < \epsilon' : |I^{\epsilon}(k) - I^{\epsilon}(k_m)| < \hat{\epsilon}/3$. Let M_2 be such that for all $l \ge M_2 : |\lim_m D^{k_m} - D^{k_l}| < \hat{\epsilon}/(3|J|)$ (recall that J > 0). Let $M = \max\{M_1, M_2\}$. Since k_M is a step function, $I^{\epsilon}(k_M)$ converges to $D^{k_M}J$ so we can find ϵ_0 such that for all $\epsilon < \epsilon_0 : |I^{\epsilon}(k_M) - D^{k_M}J| < \hat{\epsilon}/3$. Now we see for all $\epsilon < \min\{\epsilon_0, \epsilon'\}$:

$$\begin{aligned} |I^{\epsilon}(k) - \lim_{m} D^{k_{m}} J| &\leq |I^{\epsilon}(k) - I^{\epsilon}(k_{M})| + |I^{\epsilon}(k_{M}) - D^{k_{M}} J| + |D^{k_{M}} J - \lim_{m} D^{k_{m}} J| \\ &< \hat{\epsilon}/3 + \hat{\epsilon}/3 + |D^{k_{M}} - \lim_{m} D^{k_{m}} ||J| < \hat{\epsilon}/3 + \hat{\epsilon}/3 + \hat{\epsilon}/3 = \hat{\epsilon}. \end{aligned}$$

This proves that (27) holds for k and that $D^k = \lim_{m \to \infty} D^{k_m}$.

Now we can applies this to a nice class of functions.

Corollary 5.11. Let $k : [0, \infty) \to [0, \infty)$ be continuous, decreasing and compactly supported. Then (27) holds for k.

Proof. First of all, if h and h' satisfy (27) and $h' \ge h$, it is easy to see that $I^{\epsilon}(h') \ge I^{\epsilon}(h)$ for all ϵ , so $D^{h}J \ge D^{h'}J$. Since J > 0, we conclude $D^{h} \ge D^{h'}$. Now let α be such that $[0, \alpha]$ contains the support of k. Since k is continuous and compactly supported, we can approximate it with a decreasing sequence of step functions $(k_m)_{m=1}^{\infty}$ that each have support contained in $[0, \alpha]$. Since $k_m \le k_l$ for $m \ge l$, we see (by the observation above) that $D^{k_m} \le D^{k_l}$ for $m \ge l$. This means that $(D_{k_m})_{m=1}^{\infty}$ is decreasing. Since it is also clearly non-negative, it is bounded from below and hence convergent. Now proposition 5.10 tells us that (27) holds for k.

We conclude that (27) holds for all kernels that are continuous and compactly supported. This means that for all such kernels, the convergence result of this section holds (albeit with different speed constants). A simple example is $k(x) = \max\{1 - x, 0\}$ or, more generally, $k(x) = (1 - x/a)^r \mathbb{1}_{[0,a]}(x)$ for some a, r > 0.

Choosing a specific sequence

We have considered a random grid and studied the convergence of the Dirichlet forms associated to random walks on the grids as the amount of points increases to infinity and the bandwidth ϵ decreases to 0. The limit as $N \to \infty$ is an almost sure limit for every fixed f. We would like to find a fixed sequence $(p_i)_{i=1}^{\infty}$ for which the first limit holds for any continuous f. This sequence could serve as the grid that we have been looking for to define an interacting particle system on (note that the second limit does not depend on the grid). We first have the following lemma.

<u>Lemma 5.12</u>. Let f, g be continuous functions on M. Then for any $x, y \in M : |(f(x) - f(y))^2 - (g(x) - g(y))^2| \le 3(||f||_{\infty} + ||g||_{\infty})||f - g||_{\infty}$.

Proof. Writing out and using the triangle inequality shows

$$|(f(x) - f(y))^{2} - (g(x) - g(y))^{2}| \le |f(x)^{2} - g(x)^{2}| + 2|(f(x)f(y) - g(x)g(y)| + |f(y)^{2} - g(y)^{2}|.$$

Now we see

$$|f(x)^2 - g(x)^2| = |f(x) - g(x)| \cdot |f(x) + g(x)| \le ||f - g||_{\infty} ||f + g||_{\infty} \le ||f - g||_{\infty} (||f||_{\infty} + ||g||_{\infty}).$$

Similarly $|f(y)^2 - g(y)^2| \le ||f - g||_{\infty} (||f||_{\infty} + ||g||_{\infty}).$ Now for the last term.

$$\begin{split} |f(x)f(y) - g(x)(y)| &= |f(x)f(y) - f(x)g(y) + f(x)g(y) - g(x)g(y)| \\ &= |f(x)(f(y) - g(y)) + (f(x) - g(x))g(y)| \\ &\leq |f(x)| \cdot |f(y) - g(y)| + |g(y)| \cdot |f(x) - g(x)| \\ &\leq ||f||_{\infty} ||f - g||_{\infty} + ||g||_{\infty} ||f - g||_{\infty} = ||f - g||_{\infty} (||f||_{\infty} + ||g||_{\infty}). \end{split}$$

Combining these terms, we obtain $|(f(x) - f(y))^2 - (g(x) - g(y))^2| \le 3(||f||_{\infty} + ||g||_{\infty})||f - g||_{\infty}$.

Recall that we write:

$$U_f^{N,\epsilon} = \frac{1}{N(N-1)} \sum_{i \neq j=1}^N W_{ij}^{\epsilon} (f(P_j) - f(P_i))^2, \qquad a_f^{\epsilon} = \mathbb{E} W_{ij}^{\epsilon} (f(P) - f(Q))^2,$$

where P and Q are iid uniformly random elements of M. For a sequence $(p_i)_{i=1}^{\infty}$, denote

$$u_f^{N,\epsilon,p} = \frac{1}{N(N-1)} \sum_{i \neq j=1}^N W_{ij}^{\epsilon} (f(p_j) - f(p_i))^2.$$

Now we can present the result.

<u>Lemma 5.13</u>. There exists a sequence $(p_i)_{i=1}^{\infty}$ in M such that $u_f^{N,\epsilon,p} \to a_f^{\epsilon}$ for every continuous f.

Proof. Denote the space of continuous functions on M by C(M). Since M is compact, C(M) is separable with respect to the uniform distance. Let $V = \{f_k : k \in \mathbb{N}\}$ be a dense (and countable) subset of C(M). Denote $A_g = \{U_g^{N,\epsilon} \to a_g^{\epsilon}\}$. Note that we only needed for the first limit that f is bounded, so in particular we know $U_g^{N,\epsilon} \to a_g^{\epsilon}$ a.s. for all continuous functions. Hence $\mathbb{P}(A_g) = 1$ for any continuous g. Call $A = \{U_g^{N,\epsilon} \to a_g^{\epsilon} \, \forall g \in V\}$. Then $A = \bigcap_{g \in V} A_g$. This implies that

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^C) = 1 - \mathbb{P}(\bigcup_{g \in V} A_g^C) \ge 1 - \sum_{g \in V} \mathbb{P}(A_g^C) = 1,$$

so $\mathbb{P}(A) = 1$. We see that for almost every realization of $(P_i)_{i=1}^{\infty}$, we have convergence for every $f \in V$. Call any such realization $(p_i)_{i=1}^{\infty}$. Now let f be any fixed continuous function on M and let $(g_l)_{l=1}^{\infty}$ be a sequence in V such that $||g_l - f|| \to 0$ as l goes to infinity. For any two continuous functions h_1, h_2 on M, call $C^{h_1h_2} = 3(||h_1||_{\infty} + ||h_2||_{\infty})||h_1 - h_2||_{\infty}$. Note that $C^{fg_l} \to 0$ as l goes to infinity. Now we see, using lemma 5.12:

$$\begin{aligned} |u_f^{N,\epsilon,p} - u_{g_l}^{N,\epsilon,p}| &\leq \frac{1}{N(N-1)} \sum_{i \neq j} W_{ij}^{\epsilon} |(f(p_j) - f(p_i))^2 - (g_l(p_j) - g_l(p_i))^2| \\ &\leq \frac{1}{N(N-1)} \sum_{i \neq j} ||k||_{\infty} C^{fg_l} = ||k||_{\infty} C^{fg_l} \to 0, \end{aligned}$$

so $u_f^{N,\epsilon,p} - u_{g_l}^{N,\epsilon,p}$ goes to 0 as l goes to infinity uniformly in N. We can use the same estimation for another convergence statement:

$$\begin{aligned} |a_f^{\epsilon} - a_{g_l}^{\epsilon}| &\leq \mathbb{E}W_{ij}^{\epsilon} | (f(P) - f(Q))^2 - (g_l(P) - g_l(Q))^2 |\\ &\leq \mathbb{E}||k||_{\infty} C^{fg_l} = ||k||_{\infty} C^{fg_l} \to 0, \end{aligned}$$

so $a_{g_l}^{\epsilon}$ goes to a_f^{ϵ} as $l \to \infty$.

Now fix $\delta > 0$. Let L_1 be such that for all $l \ge L_1$ and for all N: $|u_f^{N,\epsilon,p} - u_{g_l}^{N,\epsilon,p}| < \delta/3$. Let L_2 be such that for all $l \ge L_2$: $|a_f^{\epsilon} - a_{g_l}^{\epsilon}| < \delta/3$. Set $L = \max\{L_1, L_2\}$. Since $g_L \in V$, we can find $N^0 \in N$ such that for all $N \ge N^0$: $|u_{g_L}^{N,\epsilon,p} - a_{g_L}^{\epsilon}| < \delta/3$. Now we obtain that for all $N \ge N^0$:

$$\begin{aligned} |u_f^{N,\epsilon,p} - a_f^{\epsilon}| &\leq |u_f^{N,\epsilon,p} - u_{g_L}^{N,\epsilon,p}| + |u_{g_L}^{N,\epsilon,p} - a_{g_L}^{\epsilon}| + |a_{g_L}^{\epsilon} - a_g^{\epsilon}| \\ &< \delta/3 + \delta/3 + \delta/3 = \delta. \end{aligned}$$

We conclude that for the sequence $(p_i)_{i=1}^{\infty}$: $u_f^{N,\epsilon,p} \to a_f^{\epsilon}$ for every continuous f.

We summarize our generalized findings with a fixed sequence in the following proposition.

<u>Proposition 5.14</u>. There is a sequence $(p_i)_{i=1}^{\infty}$ of points in M such that for any continuous, compactly supported kernel k and for any smooth function $\phi : M \to \mathbb{R}$ there is a constant C such that:

$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \epsilon^{-2-n} \frac{1}{N^2} \sum_{i,j=1}^{N} W_{ij}^{\epsilon} (\phi(p_j) - \phi(p_i))^2 = C \int_M \langle \nabla \phi, \nabla \phi \rangle \, \mathrm{d}V.$$
(28)

Moreover, the set of sequences $(p_i)_{i=1}^{\infty}$ that satisfy the above has probability 1 with respect to the product of uniform measures on M.

Note that the second statement follows from the proof of lemma 5.13, since we picked any sequence from a set of probability 1.

5.3 Convergence of the generators

Aim

Now that we have studied the convergence of the Dirichlet forms, it is easier to study the generators. However, two difficulties arise when trying to copy the strategy for Dirichlet forms to the generators. The first one is that the generator of a jumping process on $V_N = \{p_1, .., p_N\}$ is only defined for functions $V_N \to \mathbb{R}$. This makes it harder to give meaning to statements like $L_K f \to L f$ (where L_K is some generator on the functions on V_K and L is a generator on some set of functions on M). The second one is that for convergence of the generators we need uniform convergence of the generators applied to a fixed function. This uniform convergence is hard to obtain using the methods from the previous section.

We do not need convergence of the generators, but only a slightly weaker result for section 6. This solves the first difficulty. It also weakens the second difficulty, but there still remains a problem. This will become clearer in section 6.3. What we will prove now is the following statement.

Proposition 5.15. There is a sequence $(p_i)_{i=1}^{\infty}$ of points in M such that for any continuous, compactly supported kernel k and for any smooth function $\phi : M \to \mathbb{R}$ there is a constant \hat{C} such that for any i:

$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \epsilon^{-2-n} \frac{1}{N} \sum_{j=1}^{N} W_{ij}^{\epsilon}(\phi(p_j) - \phi(p_i)) = \hat{C} \Delta_M \phi(p_i).$$
(29)

Moreover, the set of sequences $(p_i)_{i=1}^{\infty}$ that satisfy the above has probability 1 with respect to the product of uniform measures on M.

Note that for each N we consider a random walk on a finite graph (with iid exponential waiting times), so we have seen before (in section 2.1) that the expression in the proposition is indeed the (rescaled) generator of the process.

Approximating integrals

We start with the following lemma.

<u>Lemma 5.16</u>. Let $(P_i)_{i=1}^{\infty}$ be an iid sequence of uniformly random elements of M. Then

$$\mathbb{P}\left(\forall f \in C(M) : \frac{1}{N} \sum_{i=1}^{N} f(P_i) \to \int_M f \mathrm{d}\bar{V}\right) = 1.$$

Proof. Let f be a continuous function. Then $\mathbb{E}(f(P_i)) = \int_M f d\bar{V} < \infty$ (recall that \bar{V} is the uniform distribution on M). Hence by the strong law of large numbers:

$$\mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N}f(P_i)\to \int_M f\mathrm{d}\bar{V}\right)=1.$$

Now let W be a countable and dense subset of C(M). By a standard argument, the above implies that

$$\mathbb{P}\left(\forall f \in W : \frac{1}{N} \sum_{i=1}^{N} f(P_i) \to \int_M f \mathrm{d}\bar{V}\right) = 1.$$

Now let $(p_i)_{i=1}^{\infty}$ be any sequence such that

$$\forall f \in W : \frac{1}{N} \sum_{i=1}^{N} f(p_i) \to \int_M f \mathrm{d}\bar{V}$$

and let f be any continuous function. Let $\hat{\epsilon} > 0$ be fixed. Since W is dense, there is a $g \in W$ such that $||g - f||_{\infty} < \hat{\epsilon}/3$. Let $L \in \mathbb{N}$ be such that for all $N \ge L$:

$$\left|\frac{1}{N}\sum_{i=1}^{N}g(p_i) - \int_{M}g\mathrm{d}\bar{V}\right| < \hat{\epsilon}/3.$$

Now we see for all $N \ge L$:

$$\begin{aligned} \left| \frac{1}{N} \sum_{i=1}^{N} f(p_i) - \int_M f \mathrm{d}\bar{V} \right| &\leq \left| \frac{1}{N} \sum_{i=1}^{N} f(p_i) - \frac{1}{N} \sum_{i=1}^{N} g(p_i) \right| + \left| \frac{1}{N} \sum_{i=1}^{N} g(p_i) - \int_M g \mathrm{d}\bar{V} \right| \\ &+ \left| \int_M g \mathrm{d}\bar{V} - \int_M f \mathrm{d}\bar{V} \right| \\ &< \frac{1}{N} \sum_{i=1}^{N} |f(p_i) - g(p_i)| + \hat{\epsilon}/3 + \int_M |g - f| \mathrm{d}\bar{V} \\ &\leq ||g - f||_{\infty} + \hat{\epsilon}/3 + ||g - f||_{\infty} < \hat{\epsilon}/3 + \hat{\epsilon}/3 + \hat{\epsilon}/3 = \hat{\epsilon}. \end{aligned}$$

The first limit

Using the lemma above, we can prove the following.

<u>Lemma 5.17</u>. There is a sequence $(p_i)_{i=1}^{\infty}$ such that for every *i* and for every continuous *f*:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} W_{ij}^{\epsilon}(f(p_j) - f(p_i)) = \int k(d(p, p_i)/\epsilon)(f(p) - f(p_i))\bar{V}(\mathrm{d}p).$$

Proof. Let $(p_i)_{i=1}^{\infty}$ be any sequence from the probability one set of lemma 5.16. Define g^i : $M \to \mathbb{R}$ by $g^i(p) = k(d(p, p_i)/\epsilon)(f(p) - f(p_i))$. Is is easy to see that g^i is continuous. Hence, by lemma 5.16, we see

$$\frac{1}{N}\sum_{j=1}^{N}W_{ij}^{\epsilon}(f(p_j) - f(p_i)) = \frac{1}{N}\sum_{j=1}^{N}g^i(p_j) \to \int_{M}g^i d\bar{V} = \int k(d(p,p_i)/\epsilon)(f(p) - f(p_i))\bar{V}(dp),$$

where we took the limit of $N \to \infty$ in the second step.

The second limit

Fix a sequence $(p_i)_{i=1}^{\infty}$ such that the first limit holds and assume that f is smooth. Now we obtain in exactly the same way as in section 5.2 for $k = \mathbb{1}_{[0,1]}$:

$$\int k(d(p,p_i)/\epsilon)(f(p) - f(p_i))\bar{V}(\mathrm{d}p) = \int_{B_p(0,1)} (f(p_i(\epsilon,\eta)) - f(p_i))\bar{V} \circ \exp \circ \lambda_\epsilon(\mathrm{d}\eta),$$

where $V \circ \exp \circ \lambda_{\epsilon}$ equals ϵ^n times some canonical μ plus a rest term μ_R . Now we can use the theory of section 4.1 to obtain in exactly the same way as in the proof of prop 5.8 that

$$\lim_{\epsilon \downarrow 0} \epsilon^{-2-n} \int k(d(p,p_i)/\epsilon) (f(p) - f(p_i)) \bar{V}(\mathrm{d}p) = \hat{C} \Delta_M f(p_i),$$

for some constant \hat{C} . Note the important fact that this constant is the same for every *i*. The reason is that μ does not depend on p_i . μ_R does, but the proof of prop 5.8 shows that μ_R vanishes in the limit. Moreover, it is easy to see from the proofs that \hat{C} only differs from C from prop 5.14 by a factor V(M) (the factor that is necessary to transform \bar{V} to V in the proof of prop 5.8). Now, in the same way as in section 5.2, this result can be extended to the case where k is continuous and compactly supported (and, of course, decreasing). We conclude that the first statement of prop 5.15 holds. Since the sequence that we used was just any sequence from a probability 1 set, the set of sequences $(p_i)_{i=1}^{\infty}$ that satisfy the first statement has probability 1, which proves the second statement of the proposition.

Combination of results

When we combine lemma 5.16, prop 5.14 and prop 5.15, we obtain the following.

Corollary 5.18. There exists a sequence $(p_i)_{i=1}^{\infty}$ in M such that for any continuous and compactly supported kernel k and for any smooth ϕ the following hold.

• There is a constant C such that:

$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \epsilon^{-2-n} \frac{1}{N^2} \sum_{i,j=1}^{N} W_{ij}^{\epsilon} (\phi(p_j) - \phi(p_i))^2 = C \int_M \langle \nabla \phi, \nabla \phi \rangle \, \mathrm{d}V.$$
(30)

• There is a constant \hat{C} such that for every i:

$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \epsilon^{-2-n} \frac{1}{N} \sum_{j=1}^{N} W_{ij}^{\epsilon}(\phi(p_j) - \phi(p_i)) = \hat{C} \Delta_M \phi(p_i).$$
(31)

• For any $g \in C(M)$:

$$\frac{1}{N}\sum_{i=1}^{N}g(p_i)\to \int_M g\mathrm{d}\bar{V}\qquad (N\to\infty).$$

Moreover, the set of sequences that satisfy the above has probability 1 (with respect to the product of uniform distributions). Further, C and \hat{C} only differ by a factor V(M).

The reason is that for every requirement the set of sequences that satisfy it has probability 1, so the set of sequences that satisfy all three of them has probability 1 too.

6 Hydrodynamic limit of the SEP

In section 5, we saw that it is possible to define a grid on a compact and connected Riemannian manifold M. We can define a random walk on the grid and when we take limits, the Dirichlet forms of the random walks converge to the Dirichlet form of Brownian motion and the generators converge pointwise to the generator of Brownian motion. In this section we will use this grid on M and these results to define an interacting particle system of which the macroscopic behaviour is the heat equation. We will first define the Symmetric Exclusion Process (SEP). Then we will discuss what a hydrodynamical limit is, in particular in the case of the SEP. After that we will prove that the hydrodynamical limit of the SEP is a solution of the weak heat equation. The main line of the proof is finished, but some details still need to be filled in. The assumptions that we need to make are mentioned in section 6.3 and we will comment on them in the concluding section.

6.1 Symmetric Exclusion Process

The Symmetric Exclusion Process (SEP) is an interacting particle system that was introduced in [20] and studied in detail in [14, Chapter 8]. The idea is that there is some (possibly countably infinite) amount of particles on a (possibly countably infinite) graph G. The particles are considered identical. Each particle jumps after independent exponential times with parameter 1 from x to y with probability p(x, y), unless the place that it wants to jump to is already occupied. In that case, the particle stays at its place. We assume that p(x, y) = p(y, x). Let $\eta_t \in \{0, 1\}^G$ denote the configuration of the particles at time t, i.e. $\eta_t(x) = 1$ if there is a particle at place $x \in G$ at time t and 0 else. We will sometimes write $\eta(p, t) = \eta_t(p)$. For any configuration η and points x, y define η^{xy} by

$$\eta^{xy}(z) = \begin{cases} \eta(x) & \text{if } z = y \\ \eta(y) & \text{if } z = x \\ \eta(z) & \text{if } z \neq x, y \end{cases}$$

An equivalent description of this process is the following. All edges (xy) have independent exponential clocks with rate p(x, y) = p(y, x). Whenever a clock rings, the particles that are at either side of the corresponding edge, jump along the edge. This means that if there are no particles, nothing happens. If there is one particle, it jumps. If there are two particle, they switch places. Since we are not interested in individual particles, the configuration stays the same in the latter case. Note that in this way there can never be more than two particles at the same place. Using the notation introduced above, we see that the generator of this process is

$$Lf(\eta) = \frac{1}{2} \sum_{x,y} p(x,y) (f(\eta^{xy}) - f(\eta)).$$

The factor $\frac{1}{2}$ is there since we count every edge twice. The domain of this generator is the set of all continuous functions $\{0,1\}^G \to \mathbb{R}$.

6.2 Hydrodynamic limit

We have already studied limits of sequences of processes in section 4 and 5. In section 4 we considered a single particle that jumps according to some distribution. The way it jumps changes with some parameter N that goes to infinity. We saw that we could obtain Brownian motion as a limiting process of random walks. In section 5 not only the dynamics of the particle varied with some parameter, but also the grid that the particle moves on. In this section we will study limits of stochastic processes in yet a different way. Not only the grid will vary but also the amount of particles that move along the grid. We will not be interested anymore in individual particles and there positions, but rather in the distribution of particle mass.

The main goal will be to show that the limiting densities satisfy a particular partial differential equation: the heat equation. This is the idea of a hydrodynamic limit. At a microscopic scale, the particles are just random walkers with some interaction, but at the macroscopic scale (where limits are taken in space and time), the behaviour is deterministic: it is described by a partial differential equation. We will use this subsection to give the basic definitions.

Radon measures

We start by introducing Radon measures, since we will use these to describe particle configurations.

<u>Definition 6.1</u>. A *Radon measure* is a Borel measure μ on a Hausdorff topological space that satisfies the following properties.

- μ is locally finite, meaning that every point has a neighbourhood with finite measure.
- μ is inner regular, i.e. for any measurable set $V: \mu(V) = \sup\{\mu(K) : K \subset V \text{ compact}\}$.

Since M is a metric space, it is Hausdorff, so we can define Radon measures on it. Denote the space of Radon measures on M by R(M). Denote $\mu(f) = \int_M f d\mu$ for any $f : M \to \mathbb{R}$ and $\mu \in R(M)$ for which $\int_M f d\mu$ is defined. The vague (or weak) topology on R(M) is defined as follows.

<u>Definition 6.2</u>. We say that a sequence of measures $(\mu_N)_{N=1}^{\infty}$ in R(M) converges vaguely or weakly to $\mu \in R(M)$ as N goes to infinity if for any continuous ϕ :

$$\mu_N(\phi) \to \mu(\phi) \text{ as } N \to \infty.$$

In general, we would need to restrict to continuous and bounded ϕ , but since M is compact, any continuous function is bounded. Since M is a Polish space (i.e. complete and separable), it can be shown that R(M) with the vague topology is a Polish space too.

Path space

Now let $D = D([0, \infty), R(M))$ denote the space of all paths $\gamma : [0, \infty) \to R(M)$ such that γ is right continuous and has left limits. On this space we can define the Skohorod metric (see for instance [18, Appendix A.2.2]). Since R(M) is a Polish space, it can be shown that D with the Skohorod metric is a Polish space too.

The process

Fix a kernel k that is continuous and has compact support. Now fix a sequence $(p_i)_{i=1}^{\infty}$ as in corollary 5.18. This will be the grid. Define the SEP $\eta^{N,\epsilon} = (\eta_t^{N,\epsilon})_{t\geq 0}$ on $p_1, ..., p_N$ through the generator

$$L^{N,\epsilon}h(\eta) = a(N,\epsilon) \sum_{i,j} W_{ij}^{\epsilon}(h(\eta^{ij}) - h(\eta)).$$

Here $\eta^{ij} := \eta^{p_i p_j}$ and $W_{ij}^{\epsilon} = k(d(p_i), (p_j))$. Further, we set

$$a(N,\epsilon) = \frac{1}{2\hat{C}N\epsilon^{2+n}},$$

where \hat{C} is from lemma 5.18. Recall that multiplying the generator with a constant is equivalent to speeding the process by that constant. This expression thus provides the time scale. Let $(X_i)_{i=1}^{\infty}$ be some (possibly degenerate) sequence of random variables taking values in $\{0, 1\}$. Set as the initial configuration $\eta_0^{N,\epsilon}(p_i) = X_i$. Initial conditions and trajectories of particle configurations Define

$$\mu_t^{N,\epsilon} = \frac{1}{N} \sum_{i=1}^N \delta_{p_i} \eta_t^{N,\epsilon}(p_i),$$

where δ_p is the Dirac measure which places mass 1 at $p \in M$. It puts a point mass at each particle and rescales it by the amount of possible positions, which represents the particle configuration $\eta_t^{N,\epsilon}$ at time t. In particular $\mu_t^{N,\epsilon}$ is a sub-probability measure and is in R(M). We assume that there exists a measurable function $\rho_0: M \to \mathbb{R}$ such that $0 \le \rho_0 \le 1$ and $\mu_0^{N,\epsilon}$ converges vaguely to $\rho_0 d\bar{V}$ in probability, i.e. for any continuous ϕ as $N \to \infty$ and $\epsilon \downarrow 0$:

$$\int_{M} \phi \mathrm{d}\mu_{0}^{N,\epsilon} \to \int_{M} \rho_{0} \phi \mathrm{d}\bar{V} \quad \text{in probability.}$$
(32)

The limit ρ_0 is the initial condition for the heat equation. If this is the case, we say that $\rho_0 dV$ is the density profile corresponding to the configurations $\eta_0^{N,\epsilon}$. Note that using measures here to represent the particles provides a bridge between separate particles (discrete measures) and density profiles (measures that are absolutely continuous with respect to V). We would like to show that if this initial condition is given, then at any time t the configurations $\eta_t^{N,\epsilon}$ have a corresponding density profile $\rho_t d\bar{V}$. Moreover, we want to show that $t \mapsto \rho_t$ solves the heat equation with initial condition ρ_0 .

Instead of dealing with this problem pointwise for each t, we will look at trajectories. As the particles move according to the SEP, $\gamma^{N,\epsilon} : [0,\infty) \to R(M)$ defined by $t \mapsto \mu_t^{N,\epsilon}$ is a random trajectory and hence a random element of D. It represents the positions of the particles over time. The initial configuration $X_1, ..., X_N$ and the dynamics of the SEP determine a distribution $Q^{N,\epsilon}$ on D. In this way we obtain a collection $\{Q^{N,\epsilon} : N \in \mathbb{N}, \epsilon > 0\}$ of measures on D.

Weak heat equation

Our aim is to show that the trajectory $\gamma^{N,\epsilon}$ converges in probability to a deterministic path of the form $t \mapsto \rho_t d\bar{V}$, where $t \mapsto \rho_t$ satisfies the weak heat equation. What we mean by the latter is that the limiting densities ρ_t satisfy the following equation for each smooth ϕ :

$$\int_{M} \phi \rho_t \mathrm{d}V - \int_{M} \phi \rho_0 \mathrm{d}V = \int_0^t \int_{M} \Delta_M \phi \rho_s \mathrm{d}V \mathrm{d}s$$
(33)

$$= -\int_{0}^{t} \int_{M} \langle \nabla \phi, \nabla \rho_{s} \rangle \, \mathrm{d}V \mathrm{d}s \tag{34}$$

$$= \int_0^t \int_M \phi \Delta_M \rho_s \mathrm{d}V \mathrm{d}s. \tag{35}$$

Once the first line is proved, it is known that ρ_t is smooth for any t > 0. Then the following lines are equivalent because of the definition of divergence. Note that V and \bar{V} only differ by a constant, so it does not matter which one we use in the equations above. Heuristically, the heat equation can be obtained from (35) in the following way. Taking derivatives with respect to t on both sides and interchanging it with \int_M on the left side yields:

=

$$\int_M \frac{\mathrm{d}\rho_t}{\mathrm{d}t} \phi \mathrm{d}V = \int_M \phi \Delta_M \rho_t \mathrm{d}V.$$

Now one has to argue that because this holds for every smooth ϕ and there are enough smooth functions on M, this implies that

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \Delta_M \rho$$

at any time point t, which is the heat equation.

Two variables

The situation that we have here is a bit different from usual hydrodynamic limits. The reason is the presence of a second parameter: the bandwidth ϵ . Instead of a fixed process for every configuration, we have a whole range of processes, depending on ϵ . This parameter only starts to play a role as soon as the process moves through time. In the proof of convergence in the next section, we will not take two limits. Instead we will use a sequence $(\epsilon_l, N_l)_{l=1}^{\infty}$ such that $\epsilon_l \downarrow 0$ and $N_l \to \infty$. In this way we obtain one sequence of grids and corresponding processes. Note that for the initial condition, we also take a limit in ϵ and N. However, the quantities do not depend on ϵ , since we look at time 0. This is therefore also the case in the example below.

Example of initial distribution

We can define the random variables $(X_i)_{i=1}^{\infty}$ to be independent Bernoulli random variables with $\mathbb{E}X_i = \rho_0(p_i)$ for some continuous function $\rho_0 : M \to \mathbb{R}$ with $0 \le \rho_0 \le 1$. Then we see as $N \to \infty$ (and then $\epsilon \downarrow 0$):

$$\mathbb{E}\left[\int \phi d\mu_0^{N,\epsilon}\right] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \phi(p_i)\eta_0^{N,\epsilon}(p_i)\right] = \frac{1}{N}\sum_{i=1}^N \phi(p_i)\mathbb{E}\eta_0^{N,\epsilon}(p_i)$$
$$= \frac{1}{N}\sum_{i=1}^N \phi(p_i)\rho_0(p_i) \to \int \phi\rho_0 d\bar{V},$$

since ϕ and ρ_0 are continuous and we chose $(p_i)_{i=1}^{\infty}$ as in corollary 5.18. Further,

$$\operatorname{var}\left[\int \phi d\mu_{0}^{N,\epsilon}\right] = \operatorname{var}\left[\frac{1}{N}\sum_{i=1}^{N}\phi(p_{i})\eta_{0}^{N,\epsilon}(p_{i})\right] = \frac{1}{N^{2}}\sum_{i=1}^{N}\phi(p_{i})\operatorname{var}(\eta_{0}^{N,\epsilon}(p_{i}))$$
$$= \frac{1}{N^{2}}\sum_{i=1}^{N}\phi(p_{i})\rho_{0}(p_{i})(1-\rho_{0}(p_{i})) \to 0.$$

Together this implies that (32) holds here for any continuous ϕ .

6.3 Convergence result

Main theorem

The proof of the convergence result is analogous to [18, Chapter 8]. Its core calculations are based on the following theorem. It is a standard martingale result, but it is also proved in [18]. We will formulate it in terms of our situation on a compact Riemannian manifold.

<u>Theorem 6.3</u>. Let $\{\eta_t, t \ge 0\}$ be a Feller process on a compact Riemannian manifold with generator L and semigroup S_t . For any function f such that both f and f^2 are in D(L), define

$$M_t = f(\eta_t) - f(\eta_0) - \int_0^t Lf(\eta_s) \mathrm{d}s.$$

Then M_t is a martingale with respect to the filtration $\mathscr{F}_t = \sigma\{\eta_r, r \leq t\}$. Moreover, its quadratic variation $\langle M, M \rangle_t$ equals $\int_0^t \gamma(s) ds$, where $\gamma(s) = (L(f^2) - 2fLf)(\eta_s)$.

Application of the theorem

First of all fix a smooth function ϕ on M. Define for $\eta \in \{0,1\}^{\{p_1,\dots,p_N\}}$: $f^N(\eta) = \frac{1}{N} \sum_{i=1}^N \eta(p_i) \phi(p_i) = \mu(f)$, where $\mu = \frac{1}{N} \sum_{i=1}^n \delta_i \eta(p_i)$. Note that since $L^{N,\epsilon}$ is the generator of a random walk on a the finite space of configurations, its domain consists of all functions on those configurations, so in particular f^N and $(f^N)^2$ are in it. Applying theorem 6.3 in this situation shows that $M^{N,\epsilon}$ defined by

$$M_t^{N,\epsilon} = f^N(\eta_t^{N,\epsilon}) - f^N(\eta_0^{N,\epsilon}) - \int_0^t L^{N,\epsilon} f(\eta_s^{N,\epsilon}) \mathrm{d}s$$
(36)

is a martingale with quadratic variation $\langle M^{N,\epsilon}, M^{N,\epsilon} \rangle_t = \int_0^t \gamma(s) ds$, where $\gamma(s) = (L^{N,\epsilon}(f^N)^2 - 2f^N L^{N,\epsilon} f^N)(\eta_s)$. Some basic manipulations show that

$$f(\eta^{ij}) - f(\eta) = -\frac{1}{N}(\phi(p_j) - \phi(p_i))(\eta(p_j) - \eta(p_i)).$$
(37)

Inserting definitions and leaving out some indexes (to keep everything clear) shows that the right hand side of (36) equals

$$\frac{1}{N} \sum_{i=1}^{N} \phi(p_i)(\eta_t(p_i)) - \frac{1}{N} \sum_{i=1}^{N} \phi(p_i)(\eta_0(p_i)) \\
- \left(-\int_0^t \frac{1}{N} \sum_{i,j=1}^{N} W_{ij}^{\epsilon}(\phi(p_j) - \phi(p_i))(\eta_s(p_j) - \eta_s(p_i)) ds \right) \\
= \mu_t^{N,\epsilon}(\phi) - \mu_0^{N,\epsilon}(\phi) - \int_0^t \frac{2a(N,\epsilon)}{N} \sum_{i,j=1}^{N} W_{ij}^{\epsilon}(\phi(p_j) - \phi(p_i))\eta_s(p_i) ds \\
= \mu_t^{N,\epsilon}(\phi) - \mu_0^{N,\epsilon}(\phi) - \int_0^t \frac{1}{N} \sum_{i=1}^{N} \eta_s(p_i) \left(2a(N,\epsilon) \sum_{j=1}^{N} W_{ij}^{\epsilon}(\phi(p_j) - \phi(p_i)) \right) ds. \quad (38)$$

Using convergence of the generators

By (31), we can write for any p_i :

$$2a(N,\epsilon)\sum_{j=1}^{N}W_{ij}^{\epsilon}(\phi(p_j)-\phi(p_i)) = \frac{1}{\hat{C}}\frac{1}{N\epsilon^{2+n}}\sum_{j=1}^{N}W_{ij}^{\epsilon}(\phi(p_j)-\phi(p_i)) = \Delta_M\phi(p_i) + E_{p_i}(N,\epsilon), \quad (39)$$

where $\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} E_{p_i}(N, \epsilon) = 0$. This shows that

$$\begin{split} &\int_{0}^{t} \frac{1}{N} \sum_{i=1}^{N} \eta_{s}(p_{i}) \left(2a(N,\epsilon) \sum_{j=1}^{N} W_{ij}^{\epsilon}(\phi(p_{j}) - \phi(p_{i})) \right) \mathrm{d}s \\ &= \int_{0}^{t} \frac{1}{N} \sum_{i=1}^{N} \eta_{s}(p_{i}) \left(\Delta_{M} \phi(p_{i}) + E_{p_{i}}(N,\epsilon) \right) \mathrm{d}s \\ &= \int_{0}^{t} \frac{1}{N} \sum_{i=1}^{N} \eta_{s}(p_{i}) \Delta_{M} \phi(p_{i}) \mathrm{d}s + \int_{0}^{t} \frac{1}{N} \sum_{i=1}^{N} \eta_{s}(p_{i}) E_{p_{i}}(N,\epsilon) \mathrm{d}s \\ &= \int_{0}^{t} \mu_{s}(\Delta_{M} \phi) \mathrm{d}s + \int_{0}^{t} \frac{1}{N} \sum_{i=1}^{N} \eta_{s}(p_{i}) E_{p_{i}}(N,\epsilon) \mathrm{d}s. \end{split}$$

Filling this in in (38) and (36), we obtain:

$$\mu_t^{N,\epsilon}(\phi) - \mu_0^{N,\epsilon}(\phi) - \int_0^t \mu_s^{N,\epsilon}(\Delta_M \phi) \mathrm{d}s = M_t^{N,\epsilon} + \int_0^t \frac{1}{N} \sum_{i=1}^N \eta_s^{N,\epsilon}(p_i) E_{p_i}(N,\epsilon) \mathrm{d}s,$$

so for any T > 0:

$$\sup_{0 \le t \le T} \left| \mu_t^{N,\epsilon}(\phi) - \mu_0^{N,\epsilon}(\phi) - \int_0^t \mu_s^{N,\epsilon}(\Delta_M \phi) \mathrm{d}s \right| \le \sup_{0 \le t \le T} \left| M_t^{N,\epsilon} \right| + \sup_{0 \le t \le T} \left| \int_0^t \frac{1}{N} \sum_{i=1}^N \eta_s^{N,\epsilon}(p_i) E_{p_i}(N,\epsilon) \mathrm{d}s \right|.$$

$$\tag{40}$$

We want to show that this expression converges to 0 in probability. We will deal with the terms on the right hand side separately.

Assumption on the error term

First of all

$$\begin{aligned} \left| \int_0^t \frac{1}{N} \sum_{i=1}^N \eta_s^{N,\epsilon}(p_i) E_{p_i}(N,\epsilon) \mathrm{d}s \right| &\leq \int_0^t \frac{1}{N} \sum_{i=1}^N |\eta_s^{N,\epsilon}(p_i)| |E_{p_i}(N,\epsilon)| \mathrm{d}s \leq \int_0^t \frac{1}{N} \sum_{i=1}^N |E_{p_i}(N,\epsilon)| \mathrm{d}s \\ &= t \frac{1}{N} \sum_{i=1}^N |E_{p_i}(N,\epsilon)|, \end{aligned}$$

 \mathbf{SO}

$$\sup_{0 \le t \le T} \left| \int_0^t \frac{1}{N} \sum_{i=1}^N \eta_s(p_i) E_{p_i}(N, \epsilon) \mathrm{d}s \right| \le TE(N, \epsilon)$$

where $E(N,\epsilon) = \frac{1}{N} \sum_{i=1}^{N} |E_{p_i}(N,\epsilon)|$. At this point we need to make the following assumption.

Assumption:
$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} E(N, \epsilon) = 0.$$

We know that this statement holds for every E_{p_i} but that is not enough. For instance it would suffice that both limits are uniform in the p_i 's. Then it is true for the following reason. *Conditional proof.* We know that

$$|E_{p_i}(N,\epsilon)| = \left|\frac{1}{\hat{C}}\frac{1}{N\epsilon^{2+n}}\sum_{j=1}^N W_{ij}^{\epsilon}(\phi(p_j) - \phi(p_i)) - \Delta_M \phi(p_i)\right|.$$

This means that

$$\lim_{N \to \infty} |E_{p_i}(N,\epsilon)| = \left| \frac{1}{\hat{C}} \frac{1}{\epsilon^{2+n}} \int_M k(d(p,p_i)/\epsilon)(\phi(p) - \phi(p_i))\bar{V}(\mathrm{d}p) - \Delta_M \phi(p_i) \right| =: g(p_i,\epsilon).$$

Now set $g(\epsilon) = \sup_i g(p_i, \epsilon)$. By uniformity of the limit as $N \to \infty$, we see that for all $\delta > 0$ there is an $N_{\delta} \in \mathbb{N}$ such that for all $N \ge N_{\delta}$: $0 \le |E_{p_i}(N, \epsilon)| \le g(\epsilon) + \delta$. For all such N we see

$$0 \le \frac{1}{N} \sum_{i=1}^{N} |E_{p_i}(N, \epsilon)| \le g(\epsilon) + \delta.$$

This implies that if the limit in N exists, it must be smaller than $g(\epsilon) + \delta$. This holds for any δ , so the limit must be under $g(\epsilon)$. By uniform convergence of the second limit, we see that $\lim_{\epsilon \downarrow 0} g(\epsilon) = 0$.

We believe that the uniform convergence of the second limit is true, because of compactness. We are less sure about the first limit (so in N). We will say some more about this in the discussion.

Convergence of the martingale to 0

Now for the other term. Since the trajectory $t \mapsto \mu_t^{N,\epsilon}$ is cadlag, so is $M^{N,\epsilon}$. Hence by Doobs inequality we see:

$$\mathbb{P}\left(\sup_{0\leq t\leq T} \left|M_t^{N,\epsilon}\right| > \delta\right) \leq \frac{\mathbb{E}|M_T^{N,\epsilon}|}{\delta}.$$
(41)

To show that $\mathbb{E}|M_T^{N,\epsilon}|$ goes to 0, it suffices to show that $\mathbb{E}\langle M^{N,\epsilon}, M^{N,\epsilon}\rangle_T$ goes to 0 (since then $\mathbb{E}\left[(M_T^{N,\epsilon})^2\right] = \mathbb{E}\langle M^{N,\epsilon}, M^{N,\epsilon}\rangle_T \to 0$ and hence $\mathbb{E}|M_T^{N,\epsilon}| \to 0$). This is what the following lemma tells us.

<u>Lemma 6.4</u>. For any T > 0:

$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \mathbb{E} \left\langle M^{N,\epsilon}, M^{N,\epsilon} \right\rangle_T = 0.$$

Proof. Recall that $\langle M^{N,\epsilon}, M^{N,\epsilon} \rangle_T = \int_0^T (L^{N,\epsilon}(f^N)^2 - 2f^N L^{N,\epsilon} f^N)(\eta_s) ds$. By writing out, one simply obtains

$$(L^{N,\epsilon}(f^N)^2 - 2f^N L^{N,\epsilon} f^N)(\eta) = \sum_{i,j=1}^N a(N,\epsilon) W^{\epsilon}_{ij} (f(\eta^{ij}) - f(\eta))^2.$$

Using (37), we see

$$(f(\eta^{ij}) - f(\eta))^2 \le \left(\frac{1}{N}(\phi(p_j) - \phi(p_i))(\eta(p_j) - \eta(p_i))\right)^2 \le \frac{1}{N^2}(\phi(p_j) - \phi(p_i))^2,$$

since $\eta(p_i) \in \{0, 1\}$ for all *i*. This shows that

$$0 \leq \langle M^{N,\epsilon}, M^{N,\epsilon} \rangle_{T} = \int_{0}^{T} (L^{N,\epsilon}(f^{N})^{2} - 2f^{N}L^{N,\epsilon}f^{N})(\eta_{s}) \mathrm{d}s$$

$$\leq \int_{0}^{T} \frac{a(N,\epsilon)}{N^{2}} \sum_{i,j=1}^{N} W_{ij}^{\epsilon}(\phi(p_{j}) - \phi(p_{i}))^{2} \mathrm{d}s = T \frac{a(N,\epsilon)}{N^{2}} \sum_{i,j=1}^{N} W_{ij}^{\epsilon}(\phi(p_{j}) - \phi(p_{i}))^{2}.$$

This implies that also

$$0 \leq \mathbb{E} \left\langle M^{N,\epsilon}, M^{N,\epsilon} \right\rangle_T \leq T \frac{a(N,\epsilon)}{N^2} \sum_{i,j=1}^N W_{ij}^{\epsilon} (\phi(p_j) - \phi(p_i))^2.$$

Now note that $\frac{a(N,\epsilon)}{N}$ is a constant times $N^{-2}\epsilon^{-2-n}$ so by (30)

$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \frac{a(N,\epsilon)}{N} \sum_{i,j=1}^{N} W_{ij}^{\epsilon} (\phi(p_j) - \phi(p_i))^2 = c$$

for some constant c. This implies that

$$0 \leq \lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \mathbb{E} \left\langle M^{N,\epsilon}, M^{N,\epsilon} \right\rangle_T \leq \lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \frac{T}{N} \frac{a(N,\epsilon)}{N} \sum_{i,j=1}^N W_{ij}^{\epsilon} (\phi(p_j) - \phi(p_i))^2 = 0.$$

We conclude that the right hand side of (41) goes to zero as N goes to infinity and ϵ goes to zero, so

$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \sup_{0 \le t \le T} \left| M_t^{N,\epsilon} \right| = 0 \quad \text{in probability.}$$

Convergence of the expression to 0 in probability

Combining everything above and using (40), we conclude that

$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} \sup_{0 \le t \le T} \left| \mu_t^{N,\epsilon}(\phi) - \mu_0^{N,\epsilon}(\phi) - \int_0^t \mu_s^{N,\epsilon}(\Delta_M \phi) \mathrm{d}s \right| = 0 \quad \text{in probability.}$$

In particular, for any $\delta \geq 0$, define

$$H^{\delta} = \left\{ \alpha \in D : \sup_{0 \le t < T} \left| \alpha_t(\phi) - \alpha_0(\phi) - \int_0^t \alpha_s(\Delta_M \phi) \mathrm{d}s \right| \le \delta \right\}.$$

It can be shown, as in [18, Chapter 8], that H^{δ} is closed for any $\delta > 0$. Recall that we write the distribution of $t \mapsto \mu_t^{N,\epsilon}$ as $Q^{N,\epsilon}$. Then the convergence result above implies that for any $\delta > 0$:

$$\lim_{\epsilon \downarrow 0} \lim_{N \to \infty} Q^{N,\epsilon}(H^{\delta}) = 1.$$

Limit distribution

We now need to reduce to one limit instead of two. Therefore we state the following.

Assumption: There is a sequence $(\epsilon_l, N_l)_{l=1}^{\infty}$ such that $\epsilon_l \downarrow 0, N_l \to \infty$ and the convergence results above hold as $l \to \infty$.

Write $Q^l = Q^{N_l,\epsilon_l}$. As in [18, Chapter 8], it can be shown that $(Q^l)_{l=1}^{\infty}$ is a tight sequence of measures on D. This implies that every one of its subsequences is also tight and therefore has a weakly convergent subsequence. If these all have the same limit, then it follows from a basis result in metric spaces that the sequence itself converges weakly to that limit. It therefore suffices for weak convergence of $(Q^l)_{l=1}^{\infty}$ to show that every weakly convergent subsequence of $(Q^l)_{l=1}^{\infty}$ has the same limit. Let $(Q^{l_k})_{k=1}^{\infty}$ be any weakly convergent subsequence and denote its limit by Q. Since H is closed, we know for any $\delta > 0$ that

$$Q(H^{\delta}) \ge \limsup_{k \to \infty} Q^{l_k}(H^{\delta}) = 1,$$

so $Q(H^{\delta}) = 1$. Since this holds for any $\delta > 0$, we see

$$Q(H^0) = Q\left(\bigcap_{m=1}^{\infty} H^{\frac{1}{m}}\right) = 1 - Q\left(\bigcup_{m=1}^{\infty} (H^{\frac{1}{m}})^C\right) \ge 1 - \sum_{m=1}^{\infty} Q\left(\left(H^{\frac{1}{m}}\right)^C\right) = 1.$$

This means that

$$Q\left(\alpha \in D: \sup_{0 \le t < T} \left| \alpha_t(\phi) - \alpha_0(\phi) - \int_0^t \alpha_s(\Delta_M \phi) ds \right| = 0 \right) = 1.$$

By doing this for a countable set of functions ϕ that is dense in C^{∞} and arguing that this implies the same for any smooth function we see:

$$Q\left(\alpha \in D: \sup_{0 \le t < T} \left| \alpha_t(\phi) - \alpha_0(\phi) - \int_0^t \alpha_s(\Delta_M \phi) \mathrm{d}s \right| = 0 \quad \forall \phi \in C^\infty \right) = 1.$$

Since this holds for any T > 0, we see that Q-a.s. for every $t \ge 0$ and for all smooth ϕ :

$$\alpha_t(\phi) - \alpha_0(\phi) = \int_0^t \alpha_s(\Delta_M \phi) \mathrm{d}s.$$
(42)

Uniqueness

Recall that we assumed that at time 0, $\mu_0^{N,\epsilon}$ converges to $\rho_0 d\bar{V}$. Now the author of [18] reasons as follows. He first shows some Q - a.s. boundedness condition on $\alpha \in D$. Then he uses a uniqueness theorem ([18, Thm A.28]). It states that when $\alpha \in D$ satisfies (42), the initial condition and the boundedness condition, then α is of the form $t \mapsto \rho_t(x) dx$ where $t \mapsto \rho_t$ is the solution to the heat equation with initial condition ρ_0 (note that setting $\alpha_t = \rho_t dV$ reduces (42) to (33)). The idea of the proof is that it passes from measures to functions via smooth approximations of the Dirac measures. Then it uses the uniqueness of the solution of the usual heat equation. The latter is a standard result in our case (so for a compact and connected Riemannian manifold) and can be found in [7], for instance. Therefore we believe that the analogous result holds in our case.

Conclusion

Now let $t \mapsto \rho_t$ be the solution to the heat equation on M with initial condition ρ_0 and call $\beta := (t \mapsto \rho_t d\bar{V})$. Recall that (42) holds Q-a.s. By the uniqueness result above, this implies that Q is a Dirac distribution with β as its support. Since this does not depend on Q^{l_k} , it must be the same for any convergent subsequence, so with arguments given above, we conclude that $Q^l \to Q$ weakly. Let γ^l denote the random trajectory $t \mapsto \mu_t^{N_l, \epsilon_l}$. Since Q is degenerate, the weak convergence implies convergence in probability, so $\gamma^l \to \beta$ in probability. This is what we wanted to show.

Summary and concluding remarks

This report consists of two main parts. Section 2 and 3 and the appendix contain background material: introductions to all the concepts that we used. These are mainly Markov processes with their generators, Dirichlet forms and Riemannian manifolds. Section 4, 5 and 6 contain our own research (although some results of section 4 were already known). We discuss these sections separately. In all of them we work with a compact and connected Riemannian manifold.

Invariance principle

In section 4 we gave a direct proof of the invariance principle. We showed uniform convergence of the generators of random walks applied to fixed smooth functions. Using Trotter-Kurtz, this implies the convergence (in distribution in the path space) of the corresponding processes. This result is already known from [10] and [2], but the proofs that are given there do not give so much insight in our situation (more general or very sketchy in a slightly different setting). That is why it was good to give a more direct proof, to see more clearly why and under which conditions the result is true.

Idea of the proofs and suitable jumping distributions

Throughout the calculations we obtained the restrictions that we need on the jumping distributions. First of all they need to be compactly supported, so the calculations stay within one coordinate chart. This needs to be, in some sense, uniform over the points of M to obtain uniform convergence. The proof with the generators then mainly consists of a Taylor approximation. To avoid that the first term blows up, we need the 'mean 0' restriction on the jumping distributions. Then to get the right factors for the second term, we need a restriction on the 'covariance matrix'. To uniformly bound the error, we locally use continuity on a compact set. This implies global uniform convergence by covering the manifold with finitely many charts.

The proof in the case of the (formal) Dirichlet forms is similar. Then we showed in section 4.3 which measures satisfy the restrictions that were obtained. We introduced canonical measures and showed that compactly supported canonical jumping distributions give convergence to a constant times Brownian motion.

Remarks on the Dirichlet form

We expected that the Dirichlet form could be an easier object to work with than the generator. This turned out to be the case in some parts of the report. However, a clear disadvantage of the Dirichlet form in the setting of a manifold is that it can be a lot harder to show that the random walk process is reversible than in \mathbb{R}^n . Even then, it is not always easy to see what the right expression for the Dirichlet form is (as is discussed in section 4.2). It would be good to do more research on both of these questions. Being symmetric is already quite a restriction for a process in \mathbb{R}^n , but it might turn out to be even more restrictive on a manifold.

For example, in the situation of the random walk of section 4.2, a minimum requirement in \mathbb{R}^n is iid jumping distributions. The 'identically distributed' part is not so natural in the setting of a manifold, since the distributions are defined on different tangent spaces. Since parallel transport gives a way to connect vector spaces, one can 'transport' a measure along a curve. One could define 'identically distributed' as: the measure in one point is just the measure from any other point, transported along a curve between the points. To be well-defined, it should not matter which pair of points or curve one uses. This puts very strong (symmetry-like) restrictions on the measures.

As we remarked at the end of section 4.2, it is possible to show convergence of the processes by just using Dirichlet forms. Then a stronger form of convergence (for instance Mosco convergence) is needed. We believe that this holds in our case, but it was not necessary to show it. It could be interesting to keep this in mind in cases where generator convergence is hard to show.

Removing restrictions on the jumping distributions

It could be interesting to see what happens in the case of measures that are not compactly supported. Then the problem is that the calculations are not local. However, maybe the same results can be obtained by approximating these measures with compactly supported measures. We get another interesting situation if we loosen the other restrictions on the jumping distributions. We expect that the processes then converge to other diffusion processes, such as Brownian motion with a drift (for instance [10] deals with such more general situations, but, again, with more abstract methods). Note, however, that in that case the limit process is not necessarily symmetric, so it is generally not possible to use the Dirichlet form.

Random grid

Properties of a good grid

As we explained in the introduction, our goal was to find a grid (actually, a sequence of grids) on a manifold on which we can construct an interacting particle system. Whether or not the (random) grid that we introduced in section 5 is suitable, thus depends on whether we can use it in section 6. This section shows that want to have a grid on which we can define a random walk such that the following properties hold.

- It is convenient if we can approximate the integral with respect to the volume measure by evaluating the function in points of the grid and averaging. We use this in the example for an initial distribution in section 6.2.
- The generators of the random processes must converge to a constant times the Laplace-Beltrami operator. The convergence should be stronger than pointwise, but it can be weaker than uniform. We use this in (39). The assumption that is stated after it shows how much stronger the convergence should be than just pointwise.
- The Dirichlet forms must converge (when applied to a fixed function) to the Dirichlet form of Brownian motion. We use this in lemma 6.4.

These properties on the random grid

As a candidate grid, we introduced a random grid. The idea is that we keep adding points that are independently chosen according to the uniform distribution on the manifold. On these random grids, we define random walks by choosing jumping probabilities that depend on some kernel k, a bandwidth ϵ and the distance between points. We first showed convergence of the Dirichlet forms when first the amount of points N goes to infinity and then ϵ to 0. The first limit comes down to a generalized law of large numbers result by Hoeffding. For the second limit, we first choose $k = \mathbb{1}_{[0,1]}$. Then we write the integral in the form of the stepping processes of section 4 and show that the jumping distribution consists of a canonical part plus a rest part that vanishes as ϵ goes to 0. This allows us to use the results from section 4. Then we use this to show that the same holds if k is continuous and compactly supported. Moreover, since the first limit is an almost sure limit, a lot of sequences exists that have this convergence.

Showing convergence of the generator (pointwise, we will get back to this) was done in the same way. Collecting these results, we concluded that there exist (deterministic) sequences that satisfy corollary 5.18, which basically almost gives the three properties above. In fact, the set of such sequences has probability 1 (with respect to the countable product of uniform measures). We make the following remarks.

Stronger than pointwise

First of all, we need that the convergence of the generators is stronger than pointwise, which is all we have in corollary 5.18. We showed in section 6.3 that it would suffice if both the limits are uniform in the p_i 's. For the limit of ϵ to 0, we think the uniformity can be shown, for the following reason. In the proof of lemma 5.5, we approximate the uniform measure in normal coordinates around a point. However, the approximation depends on the point, which gives us only pointwise convergence. If there is a way to make such an approximation uniform on the manifold, we have uniform convergence, which we need. We believe that such an approximation must be possible, since the manifold is compact. This compactness gives bounds for curvature, which should in some way bound the expression for the uniform measure. Uniform convergence on all of M implies uniform convergence on the p_i 's. Also because of compactness, it suffices to have uniform convergence on, say, open balls.

The limit of N to infinity is more tricky. The problem here is that the convergence result is just a law of large numbers that does not say much about the rate of convergence of a single realization. Is is then even harder to say something about uniformity.

Double limit

Further, there is the fact that we have a double limit instead of a single one. The beauty of how we define the random walk now, is that it apparently does not matter very much how we choose the jumping probabilities, as long as they depend in some good way on the distance between points. The problem this creates is that we have an extra variable: the bandwidth ϵ of the kernel. Ideally, we would like to express ϵ in N to obtain a single sequence of random walks on grids. However, it is not so clear how this can be done. It might depend on the manifold itself or maybe just on the dimension. We will get back to this issue later.

Criterion on grids: spreading evenly

Note that we do not really scale the grid (as we would do in \mathbb{R}^n), we just keep adding points to it. However, it is still true that the mesh size goes to 0 (with probability 1, this is easy to show). In fact, one could hypothesize that a good grid needs the following property (that implies that the mesh size goes to 0): for every set, the fraction of points of the grids in that set should converge to the uniform measure of that set. One could say that the grids should spread evenly over the manifold. The uniform sequence that we use is in some way the most natural candidate to satisfy this property (with probability 1, which gives us a lot of possible sequences), but certainly not the only one.

Hydrodynamic limit of the SEP

Interpretation

The result that we aimed for is to show that the hydrodynamic limit of the Symmetric Exclusion Process satisfies the heat equation. We take grids as in section 5 and define the SEP on each grid: independent random walkers except that jumps to occupied sites are blocked. At each time point, the particles determine a measure: a sum of Dirac measures in the points that have a particle, scaled by the amount of possible positions. As time evolves, this induces a random trajectory in the set of Radon measures on the manifold. We assume that at time 0 the particle distribution converges weakly in probability to an absolutely continuous distribution $\rho_0 d\bar{V}$. What we want to show is that then the whole trajectory converges in probability to a trajectory of the form $t \mapsto \rho_t d\bar{V}$ such that $t \mapsto \rho_t$ satisfies the weak heat equation, as given in (33), with initial condition ρ_0 .

We will first give a rough overview of the proof, then we will make some remarks.

Main line of the proof

Theorem 6.3 gives the main identity of the proof, namely that

$$M_t^{N,\epsilon} = f^N(\eta_t^{N,\epsilon}) - f^N(\eta_0^{N,\epsilon}) - \int_0^t L^{N,\epsilon} f(\eta_s^{N,\epsilon}) \mathrm{d}s$$

is a martingale with certain properties. Using convergence of the generators of random walks, we could show that this implies that

$$\mu_t^{N,\epsilon}(\phi) - \mu_0^{N,\epsilon}(\phi) - \int_0^t \mu_s^{N,\epsilon}(\Delta_M \phi) \mathrm{d}s = M_t^{N,\epsilon} + \int_0^t \frac{1}{N} \sum_{i=1}^N \eta_s(p_i) E_{p_i}(N,\epsilon) \mathrm{d}s.$$
(43)

We could then show that the martingale goes to 0 in probability (using convergence of the Dirichlet forms of random walks). We had to assume that the rest term (on the right hand side) also vanishes. Using these results, we could show that any limiting distribution Q on the path space should satisfy that Q-a.s.

$$\alpha_t(\phi) - \alpha_0(\phi) = \int_0^t \alpha_s(\Delta_M \phi) \mathrm{d}s,$$

which can be regarded a weak heat equation for measures. We believe that there is a uniqueness result for such an equation (with the fixed initial condition $\rho_0 d\bar{V}$).

We assume their exists a sequence such that the double limit is reduced to one limit. This gives a sequences of distributions. Using a tightness argument, we showed that all subsequences have weakly convergent subsequences. Because of the uniqueness that we just mentioned, their limits are equal, which implies that the sequence itself converges weakly to that limit distribution. Since the limit is degenerate, weak convergence implies convergence in probability of the trajectory.

The assumptions

To obtain the result, we had to make some assumptions. First of all we need the rest term of (43) to vanish. It became clear in section 6.3 that this has to do with the fact that pointwise convergence of the generators is not enough. We already made some comments about this above.

Further, we need that the double limit can be reduced to one limit. Then tightness gives us subsequences that still have the right limit. If we would use tightness of $\{Q^{N,\epsilon}, N \in \mathbb{N}, \epsilon > 0\}$ (with notation as in section 6), then a converging sequence in this set would not necessarily have the right limit. The difficulty is that there should be one sequence along which we have convergence of the Dirichlet forms and of the generators in every point of the grids. This is a countably infinite amount of limits. For each one of them it is easy to show the existence of such a sequence, but it is harder for all limits together. This shows the importance of the issue that we work with a double limit, as we mentioned in our discussion of section 5.

It could be worthwhile to fix ϵ and use our theory for a limit in N. Then we could study the process that is left and see if we can take the limit in ϵ . In this way we keep the double limit. However, we need some uniqueness result for the differential equation that we obtain, just as we needed it for the heat equation. It seems like a lot of the other parts of the proof would still hold.

We also assume (sometimes implicitly) that some proofs work as in the Euclidean case. For example tightness, unicity of the weak solution of the heat equation and closedness of the set H^{δ} from section 6.3. The reason is that their proofs do not seem to depend on the fact that the space is Euclidean. This makes us believe that analogous results can be proven analogously in the context of Riemannian manifolds.

Grid

We use the grid from section 5 to define the SEP on. As we noted when discussing section 5, the only properties of the grid that we need, are the ones that are listed there. This means that if we can show that any other grid satisfies these properties, we expect that we can make the same kind of construction with an analogous result. It could be worthwhile to further try to find such grids. Preferably one on which we can define random walks that only depend on N.

In that case we do not have the problems of a double limit.

Remark on the method

It is good to note that the method from [18] that we tried to follow has the following advantage. It reduces studying the SEP to studying a single random walk. The reason is that we apply the particle configurations to test functions. Then we obtain expressions that are just the generator or Dirichlet form of a random walk as in section 5, applied to the test function. This comes from the fact that the generator of the SEP is closely related to the generator of a random walk. All of this is reflected in the three properties of a good grid: they are only about the generators and Dirichlet forms of random walks.

Conclusion

Finally, we think that this approach is promising. Some points in the proof need to be strengthened or worked out a bit more, but overall we think that the grids that we defined and the method that we used work. This means that this could enable us to study interacting particle systems in the setting of a compact and connected Riemannian manifold. Of course there are a lot of questions that can be asked now. Some suggestions have been done about the grids, but we could also try to see what happens when we relax the assumptions on the manifold itself. A natural first step would be to let go of compactness and instead only assume completeness and the existence of a lower bound on the curvature.

References

- Martin T Barlow. Diffusions on fractals. In Lectures on probability theory and statistics, pages 1–121. Springer, 1998.
- [2] Gilles Blum. A note on the central limit theorem for geodesic random walks. Bull. Austral. Math. Soc, 30(2):169–173, 1984.
- [3] Olivier Bousquet, Olivier Chapelle, and Matthias Hein. Measure based regularization. In NIPS, pages 1221–1228, 2003.
- [4] Peter G. Doyle and J. Laurie Snell. Random walks and electric networks, July 2006. Derived from the book 'Random Walks and Electric Networks', 1984, by the Mathematical Association of America, Carus Monographs series.
- [5] Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda. Dirichlet forms and symmetric Markov processes, volume 19. Walter de Gruyter, 2010.
- [6] A Gaudillière. Condensers physics applied to markov chains. Lecture Notes for the 12th Brazilian School of Probability, 2009.
- [7] Alexander Grigoryan. *Heat kernel and analysis on manifolds*, volume 47. American Mathematical Soc., 2009.
- [8] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal* of the American statistical association, 58(301):13–30, 1963.
- [9] Kiyoshi Itô. The brownian motion and tensor fields on riemannian manifold. Proc. Int. Congr. Math., Stockholm, 2, 1962.
- [10] Erik Jørgensen. The central limit problem for geodesic random walks. Probability Theory and Related Fields, 32(1):1–64, 1975.
- [11] Olav Kallenberg. Foundations of modern probability. Springer Science & Business Media, 2006.
- [12] John M Lee. Riemannian manifolds: an introduction to curvature, volume 176. Springer Science & Business Media, 1997.
- [13] Ming Liao. Lévy processes in Lie groups, volume 162. Cambridge university press, 2004.
- [14] Thomas Liggett. Interacting particle systems, volume 276. Springer Science & Business Media, 2012.
- [15] Umberto Mosco. Composite media and asymptotic dirichlet forms. Journal of Functional Analysis, 123(2):368–421, 1994.
- [16] Frank Redig. Basic techniques in interacting particle systems, September 2014. Lecture notes for XXXIX summer school on mathematical physics, Ravello.
- [17] Arnaud Rousselle et al. Quenched invariance principle for random walks on delaunay triangulations. *Electronic Journal of Probability*, 20, 2015.
- [18] Timo Seppäläinen. Translation invariant exclusion processes, book in progress. Department of Mathematics, University of Wisconsin, 2008.
- [19] Amit Singer. From graph to manifold laplacian: The convergence rate. Applied and Computational Harmonic Analysis, 21(1):128–134, 2006.

- [20] Frank Spitzer. Interaction of markov processes. Advances in Mathematics, 5(2):246–290, 1970.
- [21] Michael Spivak. Comprehensive introduction to differential geometry. Vol. I. Publish or Perish, Inc., Berkeley, second edition, 1979.
- [22] Robert S Strichartz. Analysis of the laplacian on the complete riemannian manifold. Journal of functional analysis, 52(1):48–79, 1983.
- [23] Jan van Neerven. Applied functional analysis, January 2016. Course notes.
- [24] Zuoqin Wang. Lecture 14: Normal coordinates. *lecture notes of Riemannian Geometry course*, spring 2016.
- [25] Joseph Albert Wolf. Spaces of constant curvature, volume 96. McGraw-Hill New York, 1967.

A Introduction to differentiable manifolds

This appendix contains an introduction to differentiable manifolds. It contains everything that is needed to understand the topics that are discussed in section 3. We start with the definition of a smooth manifold. Then we introduce the tangent bundle, which is the space of all vectors tangent to the manifold. Next, we look at vector fields and discuss the question whether there is a curve that 'follows' the vectors in a given vector field. After that we define differential forms in order to integrate on a manifold.

A.1 Differentiable manifolds

Our intuition is coming from the fact that we are very familiar with \mathbb{R}^n . First year students get courses that introduce differentiation and integration in \mathbb{R}^n and everything corresponds with our geometric intuition, especially for n = 1, 2, 3. The idea of manifolds, however, is that they are only locally like \mathbb{R}^n . When dealing with manifolds, this is what we will constantly use. A lot of the things we define and even a lot of proofs are directly based on this local relation with \mathbb{R}^n . In all of this we mainly follow [21] and sometimes [12].

A.1.1 Smooth manifolds

Let us start with the definition.

<u>Definition A.1</u>. Let M be a metric space. We call M a *manifold* if every $x \in M$ has a neighbourhood U such that U is homeomorphic to \mathbb{R}^n for some $n \in \mathbb{N}$.

Recall that two topological spaces M and N are homeomorphic if there exists a bijective map $f: M \to N$ such that both f and f^{-1} are continuous. It is easy to show that it suffices to require that each point has a neighbourhood which is homeomorphic to an open subset of \mathbb{R}^n . Further note that for any $x \in M$, the n in the definition is unique. If this n is the same for every x, we say that M has dimension n or that it is an n-manifold. We will always implicitly assume this and we usually denote the dimension with the letter n. We see directly that \mathbb{R}^n itself and its open subsets are n-manifolds (since any point in an open subset has an open ball around it, which is homeomorphic to \mathbb{R}^n). A more interesting example is the unit sphere $S^2 \subset \mathbb{R}^3$. Each point on the sphere has a neighbourhood that is homeomorphic to \mathbb{R}^2 , so it is a 2-manifold (hence the '2' in S^2). (This in some way formalizes the explanation of a historical process: mankind has been convinced for ages that the earth is flat, because they could only observe it locally.) Note that the unit sphere itself is quite different from \mathbb{R}^2 . In fact, it is compact, whereas \mathbb{R}^2 obviously is not.

Differentiable structure

To speak of continuity of (real-valued) functions on M, we only need to know its topology. However, this is not enough to define differentiability. Luckily, differentiability is a local property and a manifold locally looks like \mathbb{R}^n . Suppose $x \in M$ and ϕ is a homeomorphism from U to \mathbb{R}^n for some neighbourhood U of x. It seems logical to say that a function $f: M \to \mathbb{R}$ is differentiable whenever $f \circ \phi^{-1}$ is. However, if V is another neighbourhood of x and ψ is a homeomorphism from V to \mathbb{R}^n , we would want to draw the same conclusion. So we want for any f that $f \circ \phi^{-1}$ is differentiable if and only if $f \circ \psi^{-1}$ is. Since $f \circ \phi^{-1} = (f \circ \psi^{-1}) \circ (\psi \circ \phi^{-1})$, we have to impose the restriction that $\psi \circ \phi^{-1}$ is differentiable. This is precisely what is done in the definition of differentiable manifolds, although we will even only consider C^{∞} functions.

<u>Definition A.2</u>. If $U, V \subset M$ are open subsets and $x : U \to x(U)$ and $y : V \to y(V)$ are homeomorphisms onto open subsets of \mathbb{R}^n , we call x and $y \subset C^\infty$ -related if the following maps

are C^{∞} :

$$x \circ y^{-1}: \ y(U \cap V) \to x(U \cap V)$$
$$y \circ x^{-1}: \ x(U \cap V) \to y(U \cap V).$$

We call a set \mathscr{A} consisting of homeomorphisms an *atlas* for M if the homeomorphisms are pairwise C^{∞} -related and their domains cover M. An element $(x, U) \in \mathscr{A}$ is then called a *chart* of \mathscr{A} or a *coordinate system* on U.

A sketch of the situation can be seen in figure 6. The maps are typically denoted by x or y to



Figure 6: φ_{α} and φ_{β} are C^{∞} -related if the compositions $\varphi_{\alpha} \circ \varphi_{\beta}^{-1}$ and $\varphi_{\beta} \circ \varphi_{\alpha}^{-1}$ are smooth as maps between $\varphi_{\alpha}(U_{\alpha} \cap U_{\beta})$ and $\varphi_{\beta}(U_{\alpha} \cap U_{\beta})$ (the dark areas). Source: https://ned.ipac.caltech.edu/level5/March01/Carroll3/Carroll2.html.

capture the idea that they represent local coordinates on M. When identifying U with \mathbb{R}^n via such a map, it is like we locally draw axes on U to give it an \mathbb{R}^n -like structure. The elements of M itself are usually denoted by p or q.

Any atlas \mathscr{A} can be extended by adding all the charts that are C^{∞} -related to all the charts in \mathscr{A} . It can be shown that the extra charts will also be C^{∞} -related to each other (so the new atlas is indeed an atlas, here we need that the charts in \mathscr{A} cover M). Any chart that is not added is not C^{∞} -related to some chart in \mathscr{A} so should not be added. In this sense the atlas that is obtained is the unique maximal atlas containing \mathscr{A} . An atlas is called maximal if it is equal to its maximal extension. We can now give the main definition.

<u>Definition A.3</u>. A pair (M, \mathscr{A}) is called a C^{∞} manifold, *differentiable* manifold or *smooth* manifold when M is a manifold and \mathscr{A} is a maximal atlas for M. \mathscr{A} is called the *differentiable* structure for M.

We will usually not mention \mathscr{A} explicitly. The following lemma shows a basic property of an atlas.

<u>Lemma A.4</u>. Let (M, \mathscr{A}) be a smooth manifold. Suppose (x, U) is a coordinate chart in \mathscr{A} and $V \subset U$ is open. Then (y, V) with $y = x|_V$ is a coordinate chart in \mathscr{A} .

Proof. It is given that V is open. Since x is a homeomorphism, x^{-1} is continuous, so x(V) is open. Since x is a continuous bijection, so is y. Moreover, $x^{-1}|_{x(V)}$ is the inverse of y and it is continuous (as the restriction of a continuous map). This shows that y is a homeomorphism

from V to the open set x(V), so it is a coordinate chart.

Now let (z, W) be any other chart from \mathscr{A} . Then $z \circ y^{-1} : y(V \cap W) \to z(V \cap W)$ is a restriction of $z \circ x^{-1} : x(U \cap W) \to z(U \cap W)$ (since y is a restriction of x and $y(V \cap W) = x(V \cap W) \subset x(U \cap W)$) and hence smooth. Similarly, $y^{-1} \circ z$ is smooth. This shows that z and y are C^{∞} -related. Since M is a smooth manifold, \mathscr{A} is maximal. The above shows that (y, V) is C^{∞} -related to any element of \mathscr{A} , so it must be in \mathscr{A} .

A natural mapping between two smooth manifolds should preserve the differentiable structure. We thus define the following.

<u>Definition A.5</u>. Let (M, \mathscr{A}) and (N, \mathscr{B}) be smooth manifolds. We call $f : M \to N$ a diffeomorphism if it is bijective and $y \in \mathscr{B}$ if and only if $y \circ f \in \mathscr{A}$. When such f exists we call M and N diffeomorphic.

A.1.2 Differentiation

Differentiability

In this setting we can start to define what differentiability means. As suggested above we say a function $f: M \to \mathbb{R}^n$ is differentiable at $p \in M$ if its composition with x^{-1} is differentiable for any coordinate system (x, U) around p. More generally we say the following.

<u>Definition A.6</u>. Let M, N be smooth manifolds and $f: M \to N$ a map. We call f differentiable (C^{∞}) if for any coordinate system (x, U) for M and (y, V) for N we have that

$$y \circ f \circ x^{-1} : \mathbb{R}^n \supset x(f^{-1}(V)) \to y(V) \subset \mathbb{R}^m$$

is differentiable (C^{∞}) . It is called *differentiable* (C^{∞}) at $p \in M$ if for all (x, U) and (y, V) such that $p \in U$ and $f(p) \in V$ the above holds at x(p).

We also call C^{∞} functions smooth functions. For differentiability (or smoothness) in a point p it suffices to show the above for any fixed pair of coordinate systems (x, U) and (y, V). It will then also hold for all other pairs. Indeed, suppose ϕ is another chart around p from the same atlas as x and ψ around f(p) from the same atlas as y. Then we have on some neighbourhood of $\phi(p)$

$$\psi \circ f \circ \phi^{-1} = (\psi \circ y^{-1}) \circ (y \circ f \circ x^{-1}) \circ (x \circ \phi^{-1}).$$

 $(\psi \circ y^{-1})$ and $(x \circ \phi^{-1})$ are differentiable on their domains (since they are from the same atlases, respectively) and $y \circ f \circ x^{-1}$ is differentiable at $x(p) = (x \circ \phi^{-1})(\phi(p))$, so $\psi \circ f \circ \phi^{-1}$ is differentiable at $\phi(p)$. Note that if we replace the two manifolds by \mathbb{R}^n for some n (with the identity map as the coordinate system, we will always implicitly assume this), this definition reduces to ordinary differentiability of a map $\mathbb{R}^n \to \mathbb{R}^m$.

Derivative

Now we would also like to calculate the derivative of a function $f: M \to \mathbb{R}$. We can define this locally with respect to a coordinate system.

<u>Definition A.7</u>. Let $f: M \to \mathbb{R}$ and $p \in M$. Let (x, U) be a coordinate system around p. Then we define the *partial derivative* of f with respect to coordinate i of x as

$$\frac{\partial f}{\partial x^i}(p) = D_i(f \circ x^{-1})(x(p)).$$

We will regard $\frac{\partial}{\partial x^i}|_p$ as the operator given by $f \mapsto \frac{\partial f}{\partial x^i}(p)$. When the underlying coordinate system and the point p are clearly known, we will sometimes denote $\partial_i := \frac{\partial}{\partial x^i}|_p$.

Here $D_i(f \circ x^{-1})$ denotes the derivative of the function $f \circ x^{-1} : \mathbb{R}^n \supset x(U) \to \mathbb{R}$ with respect to the *i*th variable. Note that this definition says that we map p to its coordinates in (x, U) and take the derivative of the composition $f \circ x^{-1}$ in these coordinates. Note that in the case where $M = \mathbb{R}^n$ (and x is the identity) the definition reduces to the definition of a partial derivative in the usual sense. We can obviously extend this definition to the case where f maps to \mathbb{R}^n . The derivatives with respect to different coordinate systems are related in the following way.

Proposition A.8. Let (x, U) and (y, V) be coordinate systems for M and let $f : M \to \mathbb{R}$ be differentiable. Then we have on $U \cap V$:

$$\frac{\partial f}{\partial x^i} = \sum_{j=1}^n \frac{\partial f}{\partial y^j} \frac{\partial y^j}{\partial x^i}.$$

This expression looks very much like the ordinary chain rule, which is exactly what is used to prove it.

Einstein summation

This is the first of many sums that we will encounter. The general convention (called Einstein summation) is that in any expression where an index appears both as a subscript and as a superscript, we assume that we sum over this index (from 1 to the dimension of the space). In this way we can write the above as

$$\frac{\partial f}{\partial x^i} = \frac{\partial f}{\partial y^j} \frac{\partial y^j}{\partial x^i}$$

Note that we regard j as a subscript in $\frac{\partial f}{\partial y^j}$, since it appears under the line.

Immersion

Now let $f: M \to N$ be a differentiable map. If (x, U) and (y, V) are coordinate systems around p and f(p), respectively, we can calculate the $m \times n$ -matrix

$$\left(\frac{\partial(y^i\circ f)}{\partial x^j}(p)\right)_{i=1..m,j=1..n}.$$

The rank of this matrix does not depend on x and y and it is called the rank of f at p.

<u>Definition A.9</u>. If the rank of f equals n (the dimension of M), we call f an *immersion*. Moreover, if f is a homeomorphism onto its image, it is called an *imbedding*.

We now have the following theorem.

<u>Theorem A.10</u>. Let M be a smooth manifold. If M is compact, then there exists an immersion $f: M \to \mathbb{R}^N$ for some $N \in \mathbb{N}$.

A.2 Tangent bundle

This section deals with a very important notion concerning manifolds: the tangent bundle. The idea is that in each point of the manifold, we would like to define a space of all the 'vectors' that are 'tangent' to the manifold in that point. This space will be of the same dimension as the manifold. One can for example think of S^2 , which has a plane as a tangent space in every point. This is easy to see when we regard S^2 as the unit sphere of \mathbb{R}^3 , but it is harder to give a definition that does not depend on this type of imbedding. This is also the order in which we will define the tangent space for a general smooth manifold: we will first regard the manifold as embedded in \mathbb{R}^m for some m to obtain an intuition and then we will give a more abstract definition that does not use such an imbedding. By doing that we (still) follow [21].

A.2.1 Intuitive definition using imbeddings

Line tangent to a curve

Let us start by looking at \mathbb{R}^n . The easiest form of the concept that we will describe is just a line tangent to an injective differentiable curve $c : \mathbb{R} \to \mathbb{R}^n$ at some point $t \in \mathbb{R}$. This line can be formed by following c up to time t and then continuing in the direction that you have at time t (and thus, generally, leaving c). This direction is described by the derivative at time t: $c'(t) = ((c^1)'(t), ..., (c^n)'(t))$ (assume that $c'(t) \neq 0$). So if one walks in this direction at speed one for one unit time after t, he ends up at c(t) + c'(t). The path that he walks since time t is a translated version of c'(t) with its starting point at c(t). For such a vector v which is placed at the point p, we will write v_p (or, if the context is clear, just v). So we can write c'(t)translated by c(t) as $c'(t)_{c(t)}$. Now if the speed with which one follows the curve is not the unit speed, but say speed 2, one will end up at $c(t) + 2c'(t) = [2c'(t)]_{c(t)}$ and if one would walk at negative unit speed, he ends up at $c(t) - c'(t) = [-c'(t)]_{c(t)}$. Now, denoting the set of all possible n-dimensional vectors starting at $p \in \mathbb{R}^n$ by \mathbb{R}_p^n , we see that we obtain a map

$$\varphi: \mathbb{R}_t \to \mathbb{R}^n_{c(t)} \qquad \qquad \lambda \mapsto [\lambda c'(t)]_{c(t)}$$

that tells us where we end up if we keep walking for a unit time with speed λ in the direction at time t. Note that the image $\varphi(\mathbb{R}) = c(t) + \mathbb{R}c'(t)$ is one-dimensional, it is the line tangent to c at time t, so at the point c(t) (in the usual sense). This is the tangent space at c(t).

Tangent spaces in Euclidean spaces

We saw above that the tangent space in a point to a curve is just a line and that the tangent vectors can be obtained by passing through a point with different speeds. We will now look at a more general case. Fix some smooth function $f : \mathbb{R}^n \to \mathbb{R}^m$. Then $f(\mathbb{R}^n)$ is a smooth subspace of \mathbb{R}^m . We will describe the tangent vectors to this subspace, we will need that later. Here we can also look at what it means to keep walking in the direction that you have at a certain point $p \in \mathbb{R}^n$ with a certain speed for a unit time. However, this speed is now described by a vector in \mathbb{R}^n starting at p representing the speed in each coordinate, so it is an element of \mathbb{R}_p^n . The direction in which you were going is now described by the $m \times n$ -matrix Df(p) of partial derivatives. The place where you end up is found by multiplying the direction matrix with the speed vector. The result is a vector in \mathbb{R}^m that is placed at starting point f(p), so it is in $\mathbb{R}_{f(p)}^m$. By doing this with every possible speed, we obtain a hyperplane in \mathbb{R}^m , which is the hyperplane tangent to f(p) (in the usual sense). This is the tangent space at f(p).

Application to imbedded manifolds

We could speak about a line tangent to a curve or a hyperplane tangent to $f(\mathbb{R}^n)$, because we interpreted them as imbedded in \mathbb{R}^n and \mathbb{R}^m , respectively. To define tangent spaces for arbitrary manifolds, we use the above in the following way. We start with a general manifold Mand an imbedding $i: M \to \mathbb{R}^N$ for some $N \in \mathbb{N}$. Now we consider a coordinate system (x, U)for M around $p \in M$. Then $i \circ x^{-1}$ is a mapping $\mathbb{R}^n \to \mathbb{R}^N$, so it is like f above. The rank of this map is n (since i is an imbedding), so in a point $i \circ x^{-1}(x(p))$ there is an n-dimensional hyperplane tangent to $i \circ x^{-1}(\mathbb{R}^n)$. We will denote this set by $(M, i)_p$ and regard it as the tangent space at p. It is easy to show that $(M, i)_p$ does not depend on the choice of x, but it does still depend on i. We write its union over $p \in M$ as T(M, i). This is a first definition of a tangent space. It can be shown to be the space of all tangent vectors to differentiable curves through i(M). The dependence on i can be shown to be insignificant: T(M, i) is essentially the same as T(M, j) for any imbedding j of M into \mathbb{R}^K . However, this construction still depends on an imbedding (which is a bit of a detour and we need it to exist) and we would like to have one canonical version of the tangent space, instead of a whole collection of them.

A.2.2 Vector bundles

The tangent bundle that we are about to construct is a specific case of a more general concept: a vector bundle. Since we will be dealing with these a lot later, we will give the general definition (from [12]) first.

<u>Definition A.11</u>. A smooth *n*-dimensional vector bundle consists of smooth manifolds E (the *total space*) and B (the *base space*) and a surjective smooth map $\pi : E \to B$ (the *projection*) such that:

- Each fibre $E_p := \pi^{-1}(p)$ is a vector space.
- Each $p \in B$ has a neighbourhood U and a diffeomorphism $\varphi : \pi^{-1}(U) \to U \times \mathbb{R}^k$ (a *local trivialisation*) such that $\pi_1 \circ \varphi = \pi$ on $\pi^{-1}(U)$, where π_1 is projection onto the first coordinate.
- The restriction of φ to p is a vector space isomorphism $E_p \to \{p\} \times \mathbb{R}^k$.

We will usually denote a vector bundle by its total space (simply E) or by specifying its projection map ($\pi : E \to B$). The intuition behind a vector bundle is that we associate a vector space to each point of a manifold (the base space) in a smooth way: the union over all these vector spaces is again a smooth manifold (the total space) and the projection map is smooth. When studying manifolds, we will use this construction with several kinds of vector spaces.

When we have a bundle over some manifold M, we can define a section of the bundle.

<u>Definition A.12</u>. Let $\pi : E \to B$ be a vector bundle. A section of E is a continuous map $s: B \to E$ such that $\pi \circ s = \text{id on } B$. If s is smooth, we call it a smooth section.

Note that a section is nothing more than a map that for each point p chooses a vector from its fibre. This is done in a continuous way: if two points are close, then the vectors that are chosen should not be very different.

A.2.3 Three constructions

The tangent bundle over a manifold M is a vector bundle with base space M. The fibre T_pM at p is the tangent space at p and the total space TM is the union of these tangent spaces. Figure 7 shows the intuitive idea of T_pM . There are (at least) three ways to construct the tangent bundle (we will present them shortly). They all first define a tangent space at a point and then take the union over these spaces. These constructions show several ways to think about the tangent bundle and they can be applied to any n-manifold M. For any smooth $f: M \to N$ we can define the map f_* in each of these constructions. It takes a vector from T_pM to $T_{f(p)}N$ in the following way: if you pass through p via a curve γ with tangent vector v at p, then the corresponding curve $f \circ \gamma$ in N has tangent vector f_*v at f(p). This is called a pushforward function, since the vector v is pushed from TM to TN. We will now describe the constructions.

Equivalence classes of tangent vectors

In the first way we look at pairs (x, v) where $x : U \to \mathbb{R}^n$ is a coordinate system around $p \in M$ and v is a vector in \mathbb{R}^n . We then call (x, v) and (y, w) equivalent if

$$w = D(y \circ x^{-1})(x(p))(v).$$

The corresponding equivalence classes $[x, v]_p$ form the tangent space at p and their union over p is the tangent bundle. For $f: M \to N$ we can define

$$f_*([x,v]_p) = [y, D(y \circ f \circ x^{-1})(x(p))(v)]_{f(p)},$$



Figure 7: The tangent space T_xM at $x \in M$. Note that $\xi \in M_x$ is a vector tangent to some curve γ through x. Source: https://www.researchgate.net/figure/228371900_fig1_Figure-1-Optimization-on-a-manifold-The-tangent-space-T-M-x.

where (x, U) and (y, v) are coordinates around p and f(p) respectively.

Equivalence classes of curves

The second way regards differentiable curves $c : (-\epsilon, \epsilon) \to M$ with c(0) = p. Let (x, U) be a coordinate system. We regard two curves c_1, c_2 equivalent if $x \circ c_1$ and $x \circ c_2$ have the same derivative at 0. The equivalence classes again form the tangent space at p and their union over p is the tangent bundle. This again shows that the tangent bundle is the collection of all possible tangent vectors to differentiable curves in M. Here we can define f_* as the map that takes the equivalence class of c to the one of $f \circ c$.

Point derivations

The third way is more abstract. We define a tangent vector at p to be a linear operator l on C^{∞} functions which is a point derivation, i.e. it satisfies Leibniz's rule:

$$l(fg) = f(p)l(g) + g(p)l(f).$$

The collection of all such operators is a vector space, this will be the tangent space at p. If (x, U) is a coordinate system around p, the derivation at $p(\frac{\partial}{\partial x^i}|_p = \partial_i)$ is in this space (this is easy to show using the product rule). In fact the set $(\partial_1, ..., \partial_n)$ is a base. This means that any tangent vector can be written in the form $l = l(x^i)\partial_i$. Now for $f: M \to N$, $f_*(l)$ can be defined by its action on C^{∞} functions g on N:

$$[f_*(l)](g) = l(g \circ f).$$

Tangent vector to a curve

If $c : \mathbb{R} \to M$ is a differentiable curve, we simplify notation. Since \mathbb{R} has only one coordinate, we call it t and write dt instead of ∂t . We also write

$$\frac{\mathrm{d}c}{\mathrm{d}t} := \frac{\mathrm{d}c}{\mathrm{d}t}(t) := \frac{\mathrm{d}c}{\mathrm{d}t}|_t := c_* \left(\frac{\mathrm{d}}{\mathrm{d}t}|_t\right),$$

where t represents both the coordinate system and a specific point in \mathbb{R} . This, again, is the vector tangent to c at c(t).

A.2.4 Orientation

If $v_1, ..., v_n$ and $w_1, ..., w_n$ are two bases for a vector space V and we define $A = a_{ij}$ to be the transformation matrix when passing from the former to the latter:

$$w_i = \sum_{j=1}^n a_{ji} v_j,$$

then we call the two bases equally oriented if det A > 0 and oppositely oriented if det A < 0(note that always det $A \neq 0$). Being equally oriented defines an equivalence relation on all bases for V and we denote the equivalence class of $v_1, ..., v_n$ by $[v_1, ..., v_n]$. Each of the (two) classes that we obtain in this way is called an orientation for V. The orientation corresponding to the unit vectors $e_1, ..., e_n$ of \mathbb{R}^n is called the standard orientation. A map $f: V \to W$ is called orientation preserving with respect to the orientation μ on V and ν on W if $[f(v_1), ..., f(v_n)] = \nu$ whenever $[v_1, ..., v_n] = \mu$. We can now define the orientation of a manifold.

<u>Definition A.13</u>. μ is called an *orientation* of the manifold M if μ_p is an orientation on each T_pM and we have a compatibility condition that basically says that if we compare $TM|_U$ to $U \times \mathbb{R}^n$ (for some open, connected U), then the orientation should either be preserved in every fibre or reversed in every fibre. If there exists an orientation of M, we call M orientable.

This definition is obviously not precise, but to make it more precise, we would have to look at equivalence of bundles and we will not do that here. If M is orientable, there are two possible orientations that differ by a minus sign. A well-known example of a non-orientable manifold is the Möbius strip.

A.3 Vector fields and differential equations

It is explained in section 3.1 that the analogue of walking along a straight line on a manifold is following a solution path of a certain differential equation. This subsection serves to provide some idea of how we can treat differential equations on a manifold.

Definition of a differential equation

We first need to define a vector field.

<u>Definition A.14</u>. A section s of TM is called a vector field. If s is C^{∞} , we call it a C^{∞} vector field. We denote the set of smooth vector fields on M by $\mathscr{T}(M)$.

A vector field is thus nothing more than a map that chooses a tangent vector in each point. This is done in a continuous way (even smooth in the case of a smooth vector field). The definition obviously coincides with the concept of a vector field in \mathbb{R}^n : here at each point in \mathbb{R}^n we have a vector in \mathbb{R}^n and this mapping $\mathbb{R}^n \to \mathbb{R}^n$ is continuous (or smooth in the case of a smooth vector field). Recall that in a coordinate system (x, U) any tangent vector can be written as a linear combination of the point derivations with respect to the coordinates of x:

$$\sum_{i=1}^{n} a^{i}(p) \frac{\partial}{\partial x^{i}}|_{p}$$

A vector field is continuous (smooth) if and only if in every coordinate system each a^i is continuous (smooth).

For a vector field on M, we can ask the question: given a certain starting point, is there a path through this point that follows the vectors in the vector field? The following definition makes this idea precise.

<u>Definition A.15</u>. Let X be a vector field on M and $\epsilon > 0$. We call a curve $\rho : (-\epsilon, \epsilon) \to M$ an *integral curve* with *initial condition* $\rho(0) = p$ if

$$\rho(0) = p
\frac{d\rho}{dt} = X(\rho(t)) \quad \forall t \in (-\epsilon, \epsilon).$$
(44)

We call (44) a differential equation.

Recall that $\frac{d\rho}{dt}$ denotes $\rho_*(\frac{d}{dt}|_t)$. This definition thus says that the tangent vector to the curve ρ at each point $\rho(t) \in M$ is given by $X(\rho(t)) \in T_{\rho(t)}M$. The question is now what the conditions on X (and M) should be for such a curve to exist and be unique. We will not completely treat that theory here, but it is good to have an idea of how this is done.

Sketch of how to solve (44)

We try to solve this problem locally. We want a neighbourhood of p on which there is a solution. The idea now is that since this is a local statement, everything can be translated to \mathbb{R}^n via a coordinate system. This yields a differential equation in \mathbb{R}^n . So we need to solve an equation of the type

$$\begin{array}{ll} \alpha(0) & = x \\ \frac{\mathrm{d}\alpha}{\mathrm{d}t} & = f(\alpha(t)) & \forall t \in (-\epsilon, \epsilon) \end{array}$$

for some suitable ϵ . To find a solution in \mathbb{R}^n we can use a fixed point theorem. Define

$$S\alpha(t) = x + \int_0^t f(\alpha(u)) \mathrm{d}u.$$

If a curve α has $S\alpha = \alpha$, then it solves $\frac{d\alpha}{dt} = f(\alpha(t))$ and has $\alpha(0) = x$, so it satisfies the differential equation. It can be shown that the other way is true as well, so α solves the differential equation if and only if it is a fixed point of S (i.e. $S\alpha = \alpha$). Now if we impose the condition that f is Lipschitz, then S is a contraction on a suitable space of functions. Using the contraction lemma (and the fact that the function space is complete) it can be shown that there exists an ϵ and α such that the equations above hold. These results can now be translated back to M. If we require the vector field to be C^{∞} , it satisfies the Lipschitz condition and we have the existence and uniqueness of a local solution.

In some cases we can say something about a global solution², namely when X only lives on a compact subspace of M, i.e. if $\operatorname{supp}(X) = \overline{\{p \in M : X(p) \neq 0\}}$ is compact. In this case we first find for each p an open neighbourhood with a local solution. Then we select finitely many of them $V_1, ..., V_k$ that cover the support of X. On each V_i there is a solution for time in $(-\epsilon_i, \epsilon_i)$. Since there are finitely many, we can take $\epsilon = \min_i \epsilon_i$. Now from each point there exists a solution for time in $(-\epsilon, \epsilon)$ (since nothing happens outside of the support of X so outside of $\cup_i V_i$). Now it can be shown that these local solutions are compatible. In fact, it can be shown that the solutions in subsequent time intervals can be glued together to obtain a solution for $t \in \mathbb{R}$. In particular when M is compact the support of X must be compact, so there is a global solution for all time.

The description above gave a sketch of how to show existence and uniqueness of solutions to a certain differential equation on M. This was a very basic differential equation. Others can be dealt with in similar ways.

 $^{^{2}}$ To make precise what we mean by global solution, we have to regard solutions as flows, which we will not do here.

A.4 Differential forms and integration

We are used to denote an integral by

This comes from the idea that we multiply the value of f at a point by an infinitisimal time interval dt and then sum over those time intervals. When integrating on a manifold, we must be more careful to give a sensible meaning to fdt. We will call it a 1-form and it will turn out to be a function of tangent vectors. To be able to understand this idea, we will first look at a simple case. Then we will introduce the theory of differential forms. Once we have those we return to defining integration on a manifold precisely.

 $\int f \mathrm{d}t.$

A.4.1 1-forms: intuitive introduction

Consider a smooth curve $c: [0,1] \to \mathbb{R}^2$ (this example is based on [21, Chapter 8]). Suppose we want to calculate $\int_c f dx + g dy$, so the integral of f dx + g dy over the curve c. Here we interpret f dx + g dy as some force field in \mathbb{R}^2 that has component f in the direction of xand component g in the direction of y. First discretize the interval [0,1] into a partition $0 = t_0 < t_1 < ... < t_n = 1$. Then choose a point to evaluate the integrands, say the left points of the intervals: $c(t_0), ..., c(t_{n-1})$. Denote the first coordinate of c by c^1 and the second by c^2 . Then on the interval (t_{i-1}, t_i) , the curve moves $c^1(t_i) - c^1(t_{i-1})$ in the direction of the first coordinate and $c^2(t_i) - c^2(t_{i-1})$ in the direction of the second one. Multiplying the right terms with each other and summing over the intervals yields the following approximation of the integral:

$$\sum_{i=1}^{n} f(c(t_{i-1}))(c^{1}(t_{i}) - c^{1}(t_{i-1})) + g(c(t_{i-1}))(c^{2}(t_{i}) - c^{2}(t_{i-1}))$$

This makes perfect sense for a curve in \mathbb{R}^2 but on a general manifold we cannot define $c^1(t_i) - c^1(t_{i-1})$, so we need another way to write the above. Since each c^j is smooth, by the mean value theorem there are points $\alpha_i, \beta_i \in (t_{i-1}, t_i)$ such that

$$c^{1}(t_{i}) - c^{1}(t_{i-1}) = (c^{1})'(\alpha_{i})(t_{i} - t_{i-1})$$

$$c^{2}(t_{i}) - c^{2}(t_{i-1}) = (c^{2})'(\beta_{i})(t_{i} - t_{i-1}).$$

Using this, we can write the approximation above in the following way:

$$\sum_{i=1}^{n} \left[f(c(t_{i-1}))(c^1)'(\alpha_i) + g(c(t_{i-1}))(c^2)'(\beta_i) \right] (t_i - t_{i-1}).$$

Since each $(c^{j})'$ is continuous, the sum above converges to the integral

$$\int_0^1 f(c(t))(c^1)'(t) + g(c(t))(c^2)'(t) \mathrm{d}t.$$

Therefore we might as well have taken any other points than α_i and β_i and still converge to the same integral. If we set $\alpha_i = \beta_i = t_{i-1}$, we obtain

$$\sum_{i=1}^{n} \left[f(c(t_{i-1}))(c^{1})'(t_{i-1}) + g(c(t_{i-1}))(c^{2})'(t_{i-1}) \right] (t_{i} - t_{i-1})$$

$$= \sum_{i=1}^{n} \left(f(c(t_{i-1})), g(c(t_{i-1}))) \bullet c'(t_{i-1})(t_{i} - t_{i-1}) \right)$$

$$= \sum_{i=1}^{n} \omega(c(t_{i-1}))(c'(t_{i-1}))(t_{i} - t_{i-1}),$$

where $\omega(p)$ is a function that maps (u, v) to f(p)u + g(p)v. Note that $\omega(p)$ is linear and we can interpret it as a function on the tangent space $T_p\mathbb{R}^2$, so it is a linear functional on $T_p\mathbb{R}^2$. It is basically the integral of f(p)dx + g(p)dy over the (tangent) vector $(x, y)_p$. This means that in the expressions above we approximate the integrand by its value at the point $c(t_{i-1})$ and the path between $c(t_{i-1})$ and $c(t_i)$ by $(t_i - t_{i-1})$ times the tangent vector to c at $c(t_{i-1})$. If we know $\omega(p)$ for each p in \mathbb{R}^2 , we can calculate the integral of fdx + gdy over any curve in \mathbb{R}^2 . We have thus transformed the integrand fdx + gdy to a linear functional on the tangent space of \mathbb{R}^2 in each point. This is something that can be generalized to arbitrary manifolds. We will eventually define the integral over a manifold in yet another way (although closely related to and motivated by the description above), but this should serve as a motivation to study functionals on tangent vectors.

A.4.2 The cotangent bundle and 1-forms

We have already seen that each tangent space T_pM is a vector space. We can use this to apply the theory of vector spaces to tangent spaces. In particular we will replace the tangent space by several related objects. To do this, we will first study some concepts concerning vector spaces.

Dual vector space

Recall that given a vector space V (in our case of finite dimension n), we can define its dual space V^* as the set of all linear functionals on V (i.e. all linear $f: V \to \mathbb{R}$). V^* is itself a vector space and we sometimes call its elements covectors. Given a basis $v_1, ..., v_n$ for V, we can define the functionals $\varphi^1, ..., \varphi^n \in V^*$ by $\varphi^i(v_j) = \delta_i^j$. This turns out to be a basis for V^* and we will call it the dual basis (with respect to $v_1, ..., v_n$, so we will sometimes denote it by $v_1^*, ..., v_n^*$). This shows that the dimension of V^* is again n.

We can also consider the dual of the dual of $V: V^{**}$. There is a natural identification $V \to V^{**}$ that maps $v \in V$ to the functional $v^{**}: \phi \mapsto \phi(v)$. In that sense we will sometimes simply write $v \in V^{**}$. If $v_1, ..., v_n \in V$ is a basis for V, then $v_1^{**}, ..., v_n^{**}$ is the dual basis of the dual basis of $v_1, ..., v_n$ and hence a basis for V^{**} . This basis equals $v_1, ..., v_n$ (through the identification that we described above). In fact, V is isomorphic to V^{**} .

Cotangent bundle and 1-forms

We can now make the following definition.

<u>Definition A.16</u>. Let M be a manifold. Let $T_p^*M = (T_pM)^*$ denote the dual space of the tangent space at $p \in M$. We define their union T^*M over $p \in M$ to be the *cotangent bundle* of M. The projection map $\pi: T^*M \to M$ takes any vector in T_p^*M to p.

It can be shown that the cotangent bundle is a smooth vector bundle and the idea is clear: at each point the tangent space is replaced by its dual space.

<u>Definition A.17</u>. Let s be a C^{∞} section of T^*M (i.e. a C^{∞} map $M \to T^*M$ that takes $p \in M$ to a vector in T_p^*M). Then we call s a 1-form on M.

This means that a 1-form on M is a map that for each p chooses a functional on the tangent vectors at p. In particular ω above was a 1-form on \mathbb{R}^2 (under smoothness conditions on f and g). A special case of a 1-form is given by the next definition.

<u>Definition A.18</u>. Let $f: M \to \mathbb{R}$ be a C^{∞} function. We define the 1-form df for $p \in M$ and $v \in T_pM$ by

$$\mathrm{d}f(p)(v) = v(f).$$

We call df the *differential* of f.

Recall that v(f) is defined via the interpretation of v as a point derivation. Not every 1-form can be obtained as a differential of some f. However we can find a nice representation. Suppose (x, U) is a coordinate system around some $p \in M$. Then we can use the coordinate functions $x^1, ..., x^n$ to define the differentials $dx^1, ..., dx^n$ (on U). Since

$$\mathrm{d}x^i(p)\left(\frac{\partial}{\partial x^j}|_p\right) = \frac{\partial}{\partial x^j}|_p x^i = \delta^j_i,$$

we see that $dx^1(p), ..., dx^n(p)$ are the dual basis for T_p^*M with respect to the basis $\frac{\partial}{\partial x^1}|_p, ..., \frac{\partial}{\partial x^n}|_p$ for T_pM . This means that we can write any 1-form (on U) as

$$\omega(p) = \sum_{i=1}^{n} \omega_i(p) \mathrm{d}x^i(p).$$

We even have the following formula for differentials:

$$\mathrm{d}f = \sum_{i=1}^n \frac{\partial f}{\partial x^i} \mathrm{d}x^i.$$

Recall that given a function $f: M \to N$, we can push a vector in T_pM forward to a vector in $T_{f(p)}N$ using the mapping f_* . With 1-forms we can do this the other way around: a 1-form ω on N naturally defines a 1-form $f^*\omega$ on M. This is called pulling back.

<u>Definition A.19</u>. Let $f: M \to N$ and let ω be a 1-form on N. Define $f^*\omega$ by

$$f^*\omega(p)(v) = \omega(f(p))(f_*v).$$

Note that $f^*\omega$ pushes v to N using f_* and then applies ω . The resulting map is a 1-form on M. Before returning to defining integration, we analyze the higher dimensional situation.

A.4.3 Covariant and contravariant tensor fields

Tensors

We can generalize the theory above by looking at functionals on products of V with itself and even with V^* using the following definition (from [12]).

<u>Definition A.20</u>. Let V be a vector space and V^* its dual. Then we define the following. A *covariant k-tensor* is a multilinear map

$$F: V \times .. \times V \ (k \text{ times}) \rightarrow \mathbb{R}.$$

A contravariant *l*-tensor is a multilinear map

$$F: V^* \times .. \times V^* \ (l \text{ times}) \to \mathbb{R}.$$

A tensor of type $\binom{k}{l}$ is a multilinear map

$$F: V^* \times .. \times V^* \ (l \text{ times}) \ \times V \times .. \times V \ (k \text{ times}) \ \rightarrow \mathbb{R}$$

The sets of these functions are denoted with $T^k(V)$, $T_l(V)$ and $T_l^k(V)$ respectively.

Note that a function is called multilinear if it is linear in each variable. In particular $T_0^k(V) = T^k(V)$, $T_l^0(V) = T_l(V)$, $T^1(V) = V^*$ and $T_1(V) = V^{**} = V$. We also define $T^0(V) = \mathbb{R}$. We now define for $F \in T_l^k(V)$ and $G \in T_q^p(V)$ their tensor product $F \otimes G$ by

$$F \otimes G(\omega^1, ..., \omega^{l+q}, v_1, ..., v_{k+p}) = F(\omega^1, ..., \omega^l, v_1, ..., v_k) G(\omega^{l+1}, ..., \omega^{l+q}, v_{k+1}, ..., v_{k+p}).$$

The following easy to check properties hold:
- $F \otimes G \in T^{k+p}_{l+q}(V)$
- $F \otimes G \neq G \otimes F$ (in general)
- $(F \otimes G) \otimes H = F \otimes (G \otimes H).$

Basis

If we denote by $\varphi^1, ..., \varphi^n$ the dual basis for V^* with respect to the basis $v_1, ..., v_n$ for V, it can be shown that the set

$$\{v_{i_1} \otimes \ldots \otimes v_{i_l} \otimes \varphi^{j_1} \otimes \ldots \otimes \varphi^{j_k}, 1 \leq i_s, j_t \leq n \ \forall s, t\}$$

is a basis for $T_l^k(V)$. This shows that its dimension is n^{k+l} . We denote the coefficients of a tensor F with respect to this basis by $F_{j_1..j_k}^{i_1..i_l}$. Sometimes we will want to have the arguments in a different order. The coefficients of a tensor $G: V \times V^* \times V \to \mathbb{R}$ will then be denoted by $G_i^{j_k}$.

<u>Remark A.21</u>. Note that we have the following convention for indexes. The indexes of elements of $V(=V^{**})$ are written as subscripts and their components as superscripts. Indexes of elements of V^* are written as superscripts and their components as subscripts. This ensures that we can make use of the Einstein summation convention in a natural way.

Trace

<u>Definition A.22</u>. Let G be a tensor of type $\binom{k+1}{l+1}$ where $k, l \ge 0$. Let $0 \le p \le k+1$ and $0 \le q \le l+1$. Then we define the *trace* of G with respect to vector p and covector j to be the tensor of type $\binom{k}{l}$ that has coefficients

$$\operatorname{tr}(G)_{j_1..j_k}^{i_1..i_l} = \sum_m G_{j_1..m..j_k}^{i_1..m..i_l},$$

where m is placed at position q between the superscripts and position p between the subscripts.

Tensor fields

As in the one-dimensional case we can replace T_pM by $T_l^k(T_pM)$ and obtain a new bundle.

<u>Definition A.23</u>. Let M be a manifold. Define $T_l^k(TM)$ as the union over $p \in M$ of $T_l^k(T_pM)$. This is called a *tensor bundle of covariant tensors of order* k. The corresponding projection $\pi : T^k(TM) \to M$ maps vectors in $T_l^k(T_pM)$ to p. A smooth section of $T_l^k(TM)$ is called a *tensor field*.

In the case that l = 0, we call $T^k(TM)$ a bundle of covariant tensors of order k and we call a smooth section a covariant tensor field.

In the case that k = 0, we call $T_l(TM)$ a bundle of contravariant tensors of order l and we call a smooth section a contravariant tensor field.

The set of tensor fields (covariant, contravariant) is denoted by \mathscr{T}_l^k ($\mathscr{T}^k, \mathscr{T}_l$).

It can be shown that this is indeed a smooth vector bundle. Note that we can apply the trace operator pointwise to a tensor field to obtain the trace of that tensor field. In analogy with the one-dimensional case, we can define for a given smooth f a pullback function f^* .

<u>Definition A.24</u>. Let $f: M \to N$ be a C^{∞} function. Let A be a covariant tensor field of order k on N. We define f^*A by

$$f^*A(p)(v_1, .., v_k) = A(f(p))(f_*v_1, .., f_*v_k).$$

In this way f^*A is a covariant tensor field of order k on M.

A.4.4 Alternating functions and k-forms

Intuitive motivation

Instead of using all of $T^k(V)$, we will be regarding a special subset. To understand this, we need to go back to the context of integration at the beginning. As we could see there, $\omega(c(t_{i-1}))(c'(t_{i-1}))$ is the integral of $f(c(t_{i-1}))dx + g(c(t_{i-1}))dy$ over the tangent vector to c at $c(t_{i-1})$. It sums over the contributions of the integrand in each coordinate. Suppose now that we would want to integrate over a surface instead of a curve. Then we would need the integral of the integrand in a point p over the area of the parallelogram spanned by the tangent vectors at p in each direction of integration. Or more generally the integral over the generalized 'parallelogram' spanned by the tangent vectors in each of k directions of integration. We could call such an integral a generalized volume where volume is locally measured by the integrand, but keep in mind that it is generally not a volume in the sense that we are used to (only in special cases).

Here is where the multilinear functionals come into play. A functional attributing this generalized volume (which is an integral) to a set of vectors must of course be linear in each coordinate: if one vector is multiplied by a constant, the volume is. The same holds when vectors are added in one coordinate of the functional: the volumes must be added. This shows that the volume functional that we look for must be an element of $T^k(T_pM)$. However, we have an extra restriction: whenever two vectors are swapped, the orientation changes, so the volume must stay the same except for a minus sign. In other words: the volume functional must be skew-symmetric. This property is captured by the set of alternating functions. So this is the set of functions that we will consider.

Alternating functions

<u>Definition A.25</u>. Let $T \in T^k(V)$. We call T alternating if $T(v_1, ..., v_i, ..., v_j, ..., v_n) = 0$ whenever $v_i = v_j$. We denote the subset of $T^k(V)$ of alternating functions by $\Omega^k(V)$.

An easy calculation shows that a function is alternating if and only if it is skew-symmetric, i.e.:

$$T(v_1, ..., v_i, ..., v_j, ..., v_n) = -T(v_1, ..., v_j, ..., v_i, ..., v_n).$$

If $T \in T^k(W)$ is alternating and $f: V \to W$ is a linear transformation between vector spaces, then f^*T defined by $f^*T(v_1, ..., v_k) = T(f(v_1), ..., f(v_k))$ is again alternating.

Let us look at an example. It can be shown that $\Omega^n(\mathbb{R}^n)$ has dimension 1: every element of $\Omega^n(\mathbb{R}^n)$ is a scalar times the determinant function on \mathbb{R}^n . The determinant itself is the unique alternating function on $(\mathbb{R}^n)^n$ (the $n \times n$ real-valued matrices) such that it maps the unit vectors $(e_1, ..., e_n)$ (the identity matrix) to 1. This implies that for any vector space V of dimension n, $\Omega^n(V)$ has dimension 1. The determinant is in some way the most basic, normalized, version of alternating functions on $(\mathbb{R}^n)^k$. In fact, all others are somehow built up from them, we will see this later.

Alternation and wedge product

We would like to have something like a tensor product for alternating functions, but the tensor product of two alternating functions is in general not alternating. Therefore we introduce a way to obtain an alternating function from any multilinear function.

<u>Definition A.26</u>. Let $T \in T^k(V)$. Define AltT by

$$\operatorname{Alt} T(v_1, ..., v_k) = \frac{1}{k!} \sum_{\sigma} \operatorname{sgn} \sigma \cdot T(v_{\sigma(1)}, ..., v_{\sigma(k)}),$$

where we sum over all permutaions σ of the set $\{1, .., k\}$ and $\operatorname{sgn} \sigma = 1$ if σ is even and -1 if σ is odd. We call Alt*T* the *alternation* of *T*.

By construction, AltT is alternating. In fact Alt $T(v_1, ..., v_k)$ is the average value of $T \circ \sigma$ over all permutations σ of those vectors, keeping in mind the extra minus when the permutation is odd. If $v_1, ..., v_k$ are rearranged, then AltT can by construction only differ in a minus sign, precisely when the rearrangement was done according to an odd permutation. This shows that AltT is alternating (since swapping two vectors is an odd permutation). Now we are ready to define the product of alternating functions.

<u>Definition A.27</u>. For $\omega \in \Omega^k(V)$ and $\eta \in \Omega^l(V)$, we define their wedge product $\omega \wedge \eta$ by

$$\omega \wedge \eta = \frac{(k+l)!}{k!l!} \operatorname{Alt}(\omega \otimes \eta).$$

Some easy properties of \wedge are:

- \wedge is bilinear
- $f^*(\omega \wedge \eta) = f^*\omega \wedge f^*\eta$
- $\omega \wedge \eta = (-1)^{kl} \eta \wedge \omega$ (anti-commutativity)
- $(\omega \wedge \eta) \wedge \theta = \omega \wedge (\eta \wedge \theta)$

Note that for odd k we have $\omega \wedge \omega = -\omega \wedge \omega$ so it must be 0.

Basis and dimension of $\Omega^k(V)$

We can now give a better characterisation of the alternating functions on a set V.

<u>Theorem A.28</u>. Let $v_1, ..., v_n$ be a basis for V and let $v_1^*, ..., v_n^*$ be the corresponding dual basis for V^* . Then the set

$$\{v_{i_1}^* \land .. \land v_{i_k}^*, 1 \le i_1 < .. < i_k \le n\}$$

is a basis for $\Omega^k(V)$. This shows that the dimension of $\Omega^k(V)$ is $\binom{n}{k}$.

Note that n must be at least equal to k. If k > n, the only alternating function is the 0 function. This theorem implies that we can write any $\omega \in \Omega^k(V)$ as

$$\omega = \sum_{i_1,\dots,i_k} \omega_{i_1,\dots,i_k} v_{i_1}^* \wedge \dots \wedge v_{i_k}^*$$

k-forms

We can repeat the same construction as we have seen a couple of times now.

<u>Definition A.29</u>. Let M be a manifold. Define $\Omega^k(TM)$ as the union over $p \in M$ of $\Omega^k(T_pM)$. The corresponding projection $\pi : \Omega^k(TM) \to M$ maps vectors in $\Omega^k(T_pM)$ to p. A smooth section of $\Omega^k(TM)$ is called a k-form (and in general a differential form).

Note that a k-form is just an alternating covariant tensor field of order k: in each point we have an alternating function on k tangent vectors. This will be the generalized volume measure. We can define the wedge product of such differential forms pointwise, using the wedge product for alternating functions.

Alternating functions in \mathbb{R}^n : linear combinations of determinants

In the next section we will properly define integration on a manifold using differential forms. However, we will now use the same intuition of this integration as in the beginning of this section (A.4.4) to better understand k-forms in \mathbb{R}^n .

In the case of \mathbb{R}^n , we have the standard basis $e_1, ..., e_n$ and the corresponding dual basis $e_1^*, ..., e_n^*$.

It can be shown that the wedge product $e_{i_1}^* \wedge .. \wedge e_{i_k}^*$ evaluated in $v_1, .., v_k$ corresponds to the determinant of the $k \times k$ -matrix

$$\left(\begin{array}{c} v_{i_1} \\ \vdots \\ v_{i_k} \end{array}\right).$$

This shows that any alternating function in $T^k(\mathbb{R}^n)$ is a linear combination of the determinants of $k \times k$ -sub-matrices of the matrix formed by the vectors in which the function is evaluated. To understand this more intuitively, consider integrating $\omega = f dx \wedge dy + g dx \wedge dz + h dy \wedge dz$ over a surface $c : [0, 1]^2 \to \mathbb{R}^3$. We then divide the surface into surface elements. We approximate these by the product of the step sizes times the integral of $f(p) dx \wedge dy + g(p) dx \wedge dz + h(p) dy \wedge dz$ over the parallelogram formed by tangent vectors at $p = c(s_{i-1}, t_{j-1})$:

$$\sum_{i,j} \omega(c(s_{i-1}, t_{j-1})) \left(\frac{\partial c}{\partial x}(s_{i-1}, t_{j-1}), \frac{\partial c}{\partial y}(s_{i-1}, t_{j-1}) \right) (t_i - t_{i-1})(s_i - s_{i-1}).$$

Each parallelogram (two-dimensional parallelogram in a three dimensional space) has projections on the xy-plane, the xz-plane and the yz-plane (this is the same as regarding the 2×2 sub-matrices, the '2 \times 2 components' of the surface element). The integral over the surface element is then the sum of the integrals over each of these projections of the integrand corresponding to that projection (i.e. the sum of the integrals over each 2×2 component, just like it was a sum of the integrals over both 1×1 components when integrating over a curve). But these integrals are simply calculated as a constant (f(p) in the $dx \wedge dy$ case) times a determinant ($dx \wedge dy$ is the determinant in the xy-plane). In this sense the 2-form $fdx \wedge dy+gdx \wedge dz+hdy \wedge dz$ on \mathbb{R}^3 is in each point p just the weighted average over determinants of 2×2 -sub-matrices where the weights are f(p), g(p) and h(p).

A.4.5 Integration

Motivation

To define integration we once again turn to the example at the beginning of section A.4. We saw that when we calculate the integral $\int_c f dx + g dy$, the approximating sums converge to the integral

$$\int_0^1 f(c(t))(c^1)'(t) + g(c(t))(c^2)'(t) \mathrm{d}t.$$

So integrating f dx + g dy over the curve c resulted in integrating $(f \circ c)(c^1)' + (g \circ c)(c^2)' dt$ over [0, 1] (in the usual way). Now observe that

$$\begin{aligned} c^*(f dx + g dy)(p)(v_p) &= (f dx + g dy)(c(p))(c_* v_p) \\ &= f(c(p)) dx(c(p))(c_* v_p) + g(c(p)) dy(c(p))(c_* v_p) \\ &= (f \circ c)(p) c^*(dx)(p)(v_p) + (g \circ c)(p) c^*(dy)(p)(v_p), \end{aligned}$$

 \mathbf{SO}

$$c^*(f \mathrm{d} x + g \mathrm{d} y) = (f \circ c)c^*(\mathrm{d} x) + (g \circ c)c^*(\mathrm{d} y).$$

We also see

$$c^{*}(\mathrm{d}x)(p)(v_{p}) = \mathrm{d}x(c(p))(c_{*}v_{p}) = (c_{*}v_{p})(x) = v_{p}(x \circ c) = v_{p}(c^{1}).$$

If we take $v_p = \frac{d}{dt}|_p$ where t denotes the identity coordinate map on [0, 1], we get

$$c^*(\mathrm{d}x)(p)\left(\frac{\mathrm{d}}{\mathrm{d}t}|_p\right) = \frac{\mathrm{d}}{\mathrm{d}t}|_p c^1 = (c^1)'(p) = (c^1)'(p)\mathrm{d}t(p)\left(\frac{\mathrm{d}}{\mathrm{d}t}|_p\right)$$

Since $\left(\frac{\mathrm{d}}{\mathrm{d}t}|_{p}\right)$ generates $T_{p}[0,1]$, we obtain

$$c^*(\mathrm{d}x) = (c^1)'\mathrm{d}t.$$

Putting everything together (with analogous results for the second term), we get:

$$c^*(f dx + g dy) = \left[(f \circ c)(c^1)' + (g \circ c)(c^2)' \right] dt.$$

This is exactly the integrand that we came across above. This means that we can compute $\int_c f dx + g dy$ by computing

$$\int_{[0,1]} c^* (f dx + g dy) = \int_{[0,1]} \left[(f \circ c)(c^1)' + (g \circ c)(c^2)' \right] dt,$$

where we can define the integral with respect to dt over [0, 1] in the usual way. In this way we will define integration.

Integration over k-cubes

<u>Definition A.30</u>. Let ω be a k-form on $[0,1]^k$. Let h be such that $\omega = h dx^1 \wedge .. \wedge dx^k$. Then we define the *integral of* ω over $[0,1]^k$ as

$$\int_{[0,1]^k} \omega := \int_{[0,1]} h \mathrm{d} x^1 .. \mathrm{d} x^k,$$

where the latter is the usual integral in \mathbb{R}^k .

Note that such h exists, since in every point ω is a scalar times the determinant, which is $dx^1 \wedge .. \wedge dx^k$.

<u>Definition A.31</u>. We call a C^{∞} function $c: [0,1]^k \to M$ a singular k-cube in M. If ω is a k-form on M, then the integral of ω over c is defined as

$$\int_c \omega := \int_{[0,1]^k} c^* \omega.$$

We have the following important property.

Proposition A.32. Let c be a singular k-cube in M and let ω be a k-form on M. If $p: [0,1]^k \to [0,1]^k$ is bijective and has $\det p' \geq 0$, then

$$\int_c \omega = \int_{c \circ p} \omega.$$

 $c \circ p$ is called a reparametrization of c that is orientation preserving. We see that the integral remains the same when we parametrize the k-cube in a different way. If $\det p' \leq 0$, we call p orientation reversing and there will appear a minus sign in front of the integral.

Integration over M

We can now integrate over singular k-cubes in M, but we would like to define an integral over all of M. If ω is an n-form on M and $c_1, c_2 : [0, 1]^n \to M$ are two orientation preserving singular n-cubes such that $\operatorname{supp} \omega \subset c([0, 1]^n)$, then it can be shown that $\int_{c_1} \omega = \int_{c_2} \omega$. We call this number $\int_M \omega$ (note that it depends on the orientation of M, the opposite orientation gives a minus sign). We now know how to define the integral over M if the n-form has support contained in the image of a singular n-cube. We will use this to define the integral over M of any ω with compact support.

If ω is an *n*-form on M, we can find an open cover \mathcal{O} of M such that each element $U \in \mathcal{O}$ is contained in the image of a singular *n*-cube. Now there exists a partition of the unit Φ subordinate to \mathcal{O} , i.e. $\Phi = \{\Phi_i, i \in I\}$ for some index set I and:

- each Φ_i is a C^{∞} function $M \to [0, 1]$
- the collection of sets $\{p : \Phi_i(p) \neq 0\}$ is locally finite
- for each $p \in M : \sum_{i \in I} \Phi_i(p) = 1$
- for each $i \in I$ there is a $U \in \mathcal{O}$ such that $\operatorname{supp} \Phi_i \subset U$.

Note that the second property means that every p has a neighbourhood on which only finitely many Φ_i are not identically 0. In particular, the sum in the third property only contains finitely many non-0 terms. For any Φ_i , the support of $\Phi_i \omega$ is contained in the image of some singular n-cube, so $\int_M \Phi_i \omega$ is defined above. Now assume that ω has compact support. This implies that it only intersects finitely many of the sets $\{p : \Phi_i \neq 0\}$. This means that $\omega = \sum_i \Phi_i \omega$ is well defined (only finitely many non-zero terms) and true since the Φ_i sum to 1. In accordance to this we define

$$\int_M \omega = \sum_i \int_M \Phi_i \omega.$$

It can be shown that this definition is independent of the chosen partition of unity (but, of course, still dependent on the orientation of M). If the support of ω is not compact, we can still use this approach, but we need a more subtle restriction on ω .