UTRECHT UNIVERSITY

BACHELOR THESIS

Heat Equation and Brownian Motion on Riemannian Manifolds

Author:
Bart HEEMSKERK

Supervisor:
Dr. Rik VERSENDAAL,
Dr. Álvaro del PINO GOMEZ

A thesis submitted in fulfilment of the requirements for the degree of Bachelor of Science Mathematics

in the

Department of Mathematics
Faculty of Science

June 14, 2021

"It remains conceivable that the measure relations of space in the infinitely small are not in accordance with the assumptions of our geometry, and, in fact, we should have to assume that they are not if, by doing so, we should ever be enabled to explain phenomena in a more simple way."

Berhnhard Riemann

UTRECHT UNIVERSITY

Abstract

Faculty of Science
Department of Mathematics

Bachelor of Science Mathematics

Heat Equation and Brownian Motion on Riemannian Manifolds

by Bart HEEMSKERK

Riemannian geometry is used in many theoretical models such as general relativity. Theoretical physicists and mathematicians have researched this extensively to describe the behaviour of particles and interactions in spacetime. These interactions include solutions to the heat equation and Brownian motion. In this thesis, we give an elaborate theoretical description of differential geometry starting from basic topology and expand it to develop a thorough conceptualization of Riemannian geometry including geodesics and Riemannian distance. We formalize the heat equation on compact Riemannian manifolds and show the existence and uniqueness of the solution that commences as a point source called the fundamental solution and derive an explicit expression using the Euclidean case. In addition, we show that the transition density function of Brownian motion that determines the likelihood of terminating within a certain region on compact Riemannian manifolds is precisely given by the fundamental solution of the heat equation. Finally, we show some applications of these two results and give an overview of present and future research.

Acknowledgements

The past two months have been a great learning experience. During my entire mathematics career, I have never worked this hard to contribute to a project before – and I find that logical. Although learning new math through courses is fun, diving headfirst into this specific field of mathematics by myself has been more exciting than I initially imagined.
When I started working on this project, I knew I would have to learn a lot of new theory – I didn't even know how differentiable functions were defined on manifolds, or what a differentiable manifold was, for that matter. Luckily, my Daily Supervisor Rik Versendaal was there to provide some insightful literature on differential- and Riemannian geometry. Even so, this field of math is rather complicated. That is why I thank him for, amongst other things, meeting up with me (almost) every week and discussing mathematics elaborately with me with an open and comfortable ambiance, every now and then taking the time to digress and discuss some interesting pieces of physics and/or mathematics that are irrelevant to this project. Even though we met online only, I feel he has given me a peek into what it's like to work as "a mathematician". I thank him in particular for handing me constructive feedback, both on my work during the past two months as well as my written thesis. In addition, massive thanks to my Supervisor Álvaro del Pino Gomez. He provided great feedback and some very interesting discussions. To my regret, I wasn't his only Bachelor student and time restricted us. Still, huge thanks to Álvaro and Rik to make room for my ambitious plans, while guiding me not to be too ambitious since the lack of time and knowledge could constrict me as we approached the deadline.

It is with great regret that I mention that the vast majority of my thesis (and research) was done at home, by myself. I did get to work together every now and then with Richard Dirven at the university. Thanks to him for accompanying me, discussing math with me and – most importantly, of course – sharing with me this good-looking LaTeX template.

Finally, I'd like to thank Barbara van den Berg for arranging the Bachelor project presentations in person this year. I enjoyed watching the other presentations and actually be able to interact with other students after we did our talks. Too bad that there was no barbecue, but as you said, we will have a second chance during our Master's theses!

Contents

Abstract **ii**

Acknowledgements **iii**

1 Introduction **1**

I Theory **3**

2 Differential Geometry **4**
- 2.1 Differentiable Manifolds 4
- 2.2 Tangent space 5
- 2.3 Derivations 8
- 2.4 Vector Fields 14

3 Riemannian Geometry **17**
- 3.1 Bilinear forms 17
- 3.2 Riemannian metrics 19
  - 3.2.1 Choice of charts 19
  - 3.2.2 Examples of Riemannian metrics 21
- 3.3 Differentiating Vector Fields 26
- 3.4 Geodesics 32
- 3.5 Riemannian Distance 39

II Heat Equation and Brownian Motion **43**

4 Heat Equation on Riemannian Manifolds **44**
- 4.1 The heat equation 44
- 4.2 The Laplace-Beltrami operator 45
- 4.3 Heat equation on Riemannian manifolds and the Fundamental Solution 48
- 4.4 The fundamental solution on $\mathbb{R}^n$ 51
- 4.5 Fundamental solution on compact Riemannian manifolds 55

5 Brownian Motion **66**
- 5.1 Definitions 66
- 5.2 Transition density in $\mathbb{R}^n$ 70
- 5.3 Transition density on Riemannian manifolds 71

6 Applications & Outlook **76**
- 6.1 Fundamental solution and geodesic distance 76
  - 6.1.1 The problem 76
  - 6.1.2 Verified cases 77
- 6.2 Escaping Brownian motion 78
- 6.3 Short-time asymptotics 80
  - 6.3.1 Asymptotic Expansion 80

6.3.2 Distant points 82

7 Conclusion **86**

A Bibliography **89**

List of Symbols

$\mathbb{N}$	The set of natural numbers $1, 2, \ldots$
$\mathbb{Z}$	The set of integers $0, \pm 1, \pm 2, \ldots$
$\mathbb{R}$	The set of real numbers
$\mathbb{C}$	The set of complex numbers
$f\|_p$	The function f restricted to point or region p
$\in$	Is an element of
$\subset$	Is a subset of or equals
$\circ$	Composed with
$\cdot$	Multiplied by
$A \to B$	Maps from domain A to codomain B
$a \mapsto b$	Maps element a to element b
$\implies$	Implies
$\iff$, iff	If and only if
$\cong$	Is isomorphic to
$\propto$	Is proportional to
$C^k(A)$	k times continuously differentiable functions on the set A
$C^\infty(A)$	smooth functions on the set A

1. Introduction

Ever since the apple fell on the head of Isaac Newton in the 17th century, mathematicians and physicists have been working on theories to describe gravity. At the end of the 19th century it became clear that velocity affected the flow of time and that time and space could not be described seperately, giving rise to a new branch of differential geometry called Minkowski spacetime. Not long after that, in the 20th century, Minkowski's former student Albert Einstein developed the theory of general relativity and theorized that gravity curves spacetime and that spacetime could be described using *manifolds* – curved spaces that, locally, look like Euclidean spaces.

It was already in 1854 that the formalism that Einstein used to develop his theory of general relativity was developed by Bernhard Riemann. In this year, he published his work "Ueber die Hypothesen, welche der Geometrie zu Grunde liegen", that is, "On the Hypotheses on which Geometry is Based" [1]. Riemann revolutionized the field of differential geometry by laying out a mathematical framework that introduced the notion of distance, angles, length and energy of curves, surface area and volume on manifolds. The framework was based on a metric tensor which enriched manifolds to be a Riemannian manifold. This allowed physicists and mathematicians to do explicit calculations on manifolds.

One of the phenomena that theorists could describe is how heat and temperature flow throughout a Riemannian manifold. The behaviour of heat flows is governed by solutions to the *heat equation*, a partial differential equation that all functions describing diffusive processes generally solve. It was the mathematician Beltrami who generalized the heat equation to Riemannian manifolds rather than just Euclidean space.

Another example that is closely related to the heat equation is Brownian motion, which gives a mathematical description of the random movement of a particle suspended in a certain medium. In probability theory, Brownian motion is a stochastic process that admits such random movement subject to a Gaussian probability distribution. Riemannian geometry gave rise to the stochastic analysis on manifolds and mathematicians have formalized the behaviour of Brownian motion on manifolds for the past century.

In this thesis, we aim to show two important results in Riemannian geometry. The first goal is to show the intuitive fact that any compact Riemannian manifold admits a solution to the heat equation whose initial temperature distribution is entirely concentrated on a single point on the manifold, and that it is unique. This solution is called the *fundamental solution* of the heat equation. From it, we are able to calculate solutions given any (finite) initial temperature distribution.

The second objective concerns Brownian motion and the probability that a "particle" ends up in a specific region on a Riemannian manifold after random motion during a specific positive time. The probability density associated to this continuous distribution is called the *transition density function*. The main goal is to show that the transition density function of Brownian motion on Riemannian manifolds is precisely given by the fundamental solution of the manifold.

The structure of this thesis is as follows.
In Chapter 2, we introduce the reader to the concept of differential geometry assuming the reader is familiar with basic topology, analysis and calculus. We explore the differentiability and derivatives of functions on manifolds.

In Chapter 3, we expand the theory on differential geometry to Riemannian geometry and introduce the notions of Riemannian metrics, differentiability and derivatives of vector fields, energy-minimizing curves called *geodesics* and the Riemannian distance function.

In Chapter 4, we introduce the heat equation on compact Riemannian manifolds and prove the first objective after letting the Euclidean case inspire us for an ansatz. Moreover, we derive an expression for the fundamental solution on compact Riemannian manifolds.

In Chapter 5, we introduce the stochastic process Brownian motion on Riemannian manifolds and prove the second objective. Similar to the heat equation, we first take a look at the Euclidean case.

Finally, in Chapter 6, we discuss some applications of the two main statements of this thesis and elaborate on current and future research. Topics include the fading away of the fundamental solution of the heat equation as a function of Riemannian distance, the disappearing property of Brownian motion on certain manifolds and the behaviour of the fundamental solution and Brownian motion throughout manifolds in the limit of short time duration.

Part I

Theory

2. Differential Geometry

This chapter includes a short introduction to differential geometry covering the parts necessary for understanding Riemannian geometry and how the heat equation can be considered within the field. We will follow [2] closely, with additions from [3]. We assume that the reader is familiar with basic topology and concepts such as homeomorphisms and topological manifolds.

2.1 Differentiable Manifolds

Reminder. A topological space $(M,\mathscr{T})$ is an n**-dimensional topological manifold** if the following holds:

(i) $(M,\mathscr{T})$ is Hausdorff;

(ii) $(M,\mathscr{T})$ is second-countable, i.e. $(M,\mathscr{T})$ admits a countable basis;

(iii) $(M,\mathscr{T})$ is locally Euclidean. This means that for all $p \in M$ there exists an open subset $U \subset M$ such that $p \in U$ and an open subset $V \in \mathbb{R}^n$ such that U is homeomorphic to V (denoted $U \cong V$).

Remark. Often we simply write M *is an* n*-manifold*.

Definition 2.1. Let M be an n-manifold. Then the homeomorphisms $x : U \to V$ are called **charts** of M.

In topology, it is straightforward to discuss continuous functions (maps). Let M be a topological manifold and take a map $f : M \to \mathbb{R}$. Since we are talking about differential geometry, it makes sense to look at the *differentiability* of f. But what does this mean?
An intuitive approach to define this is to look at how f is differentiable in the subset of $\mathbb{R}^n$ rather than M itself. In other words, for a chart $x : U \to V$, we look at how f behaves in points in U that are linked to points in V via x. One could try the following definition:

f is called *differentiable* at $p \in M$ if, for some chart $x : U \to V$ with $p \in U$, the function $f \circ x^{-1} : V \to \mathbb{R}$ is differentiable in $x(p)$.

In essence, this definition says that f is differentiable on M if it is differentiable in $\mathbb{R}^n$ (when expressed in coordinates). However, this definition is prone to problems if $p \in M$ is in the domain of multiple charts. Let $y : \tilde{U} \to \tilde{V}$ be another chart such that $p \in \tilde{U}$. Then (near $y(p)$) we have

$$f \circ y^{-1} = (f \circ x^{-1}) \circ (x \circ y^{-1}).$$

We see that the first part $f \circ x^{-1}$ is differentiable at $x(p)$, but that the second part $x \circ y^{-1}$ is only continuous. Hence, in this definition, differentiability depends on the choice of charts. This is something we do not want, since there are no preferred coordinate systems on topological manifolds. This problem can be circumvented if $x \circ y^{-1}$ is not only a homeomorphism, but also a differentiable homeomorphism, i.e. a diffeomorphism. In that case, differentiability of $f \circ x^{-1}$ implies differentiability of $f \circ y^{-1}$. To enforce this notion, let us define the following.

Definition 2.2. Let M be an n-manifold. A pair of charts of M $x : U \to V$ and $y : \tilde{U} \to \tilde{V}$ are called C^∞**-compatible** if

$$y \circ x^{-1} : x(U \cap \tilde{U}) \to y(U \cap \tilde{U})$$

is a $C^\infty-$diffeomorphism (i.e. a homeomorphism that is infinitely continuously differentiable (and so is its inverse)).

Definition 2.3. A set of charts $x_i : U_i \to V_i,\ i \in I$ for some indexing set I is called an **atlas** of M if the U_i cover M. An atlas A is called a $C^\infty-$**atlas** if any two charts in A are $C^\infty-$compatible.

Definition 2.4. A $C^\infty-$atlas A_{max} is called **maximal** or a **differentiable structure** if a chart that is $C^\infty-$compatible with all charts in A_{max} is already contained in A_{max}.

The differentiable structure allows us to talk about a set of charts that are all compatible with each other, i.e. for which the attempted definition is reasonable, since any $x \circ y^{-1}$ is infinitely continuously differentiable. Hence we arrive at the following:

Definition 2.5. Let M be an $n-$manifold and let A_{max} be a differentiable structure on M. Then the pair (M, A_{max}) is called an $n-$**dimensional differentiable manifold**.

Now, we can define what it means for a map between differentiable manifolds to be differentiable.

Definition 2.6. Let M and N be differentiable manifolds, let $p \in M$ and let $k \in \mathbb{N} \cup \{\infty\}$. A map $f : M \to N$ is called k**-times continuously differentiable** (i.e. C^k) **near** p if for all charts $(x : U \to V) \in A_{\text{max}}(M)$ with $p \in U$ and every chart $(y : \tilde{U} \to \tilde{V}) \in A_{\text{max}}(N)$ with $f(p) \in \tilde{U}$ there exists a neighborhood $W \subset x(f^{-1}(\tilde{U}) \cap U)$ of $x(p)$ such that

$$y \circ f \circ x^{-1} : x(f^{-1}(\tilde{U}) \cap U) \to \tilde{V}$$

is C^k on W.

2.2 Tangent space

Now that we can define a differentiable map f between differentiable manifolds, we can ask ourselves how to define the *derivative* of f in $p \in M$. For maps $\mathbb{R} \to \mathbb{R}$ the derivative at a point p is given by the slope of the tangent line through p. This concept can be extended to maps with multi-variable inputs with help of partial derivatives. In any case, the derivative is a *linear approximation* of the map in a certain point. We can define the derivative of a differentiable map between differentiable manifolds similarly, employing the notion of a linear approximation as follows.

Definition 2.7. Let M be a differentiable manifold, let $p \in M$ and let $x : U \to V$ a chart with $p \in U$. A **tangent vector** on M at the point p is an equivalence class of differentiable curves $c : (-\epsilon, \epsilon) \to M$ with $\epsilon > 0$ and $c(0) = p$, where the equivalence relation on such curves is given by $c_1 : (-\epsilon_1, \epsilon_1) \to M \sim c_2 : (-\epsilon_2, \epsilon_2) \to M$ if and only if

$$\frac{d}{dt}(x \circ c_1)|_{t=0} = \frac{d}{dt}(x \circ c_2)|_{t=0}.$$

In other words, two curves are called equivalent if they have the same "slope" at p. Note that the definition of the tangent vector does not depend on the choice of chart. Let $y : \tilde{U} \to \tilde{V}$ be another chart such that $p \in \tilde{U}$. Then the chain rule tells us that

$$\frac{d}{dt}(y \circ c)|_{t=0} = \frac{d}{dt}\big((y \circ x^{-1}) \circ (x \circ c)\big)|_{t=0} = D(y \circ x^{-1})|_{x(p)}\big(\frac{d}{dt}(x \circ c)|_{t=0}\big) \tag{2.1}$$

where $D(y \circ x^{-1})$ denotes the total derivative of $y \circ x^{-1}$. This tells us that

$$\frac{d}{dt}(x \circ c_1)|_{t=0} = \frac{d}{dt}(x \circ c_2)|_{t=0} \quad \text{iff} \quad \frac{d}{dt}(y \circ c_1)|_{t=0} = \frac{d}{dt}(y \circ c_2)|_{t=0}.$$

Notation. The equivalence class of c is denoted c^0.

We have introduced the notion of a linear approximation as tangent vectors. However, similar to maps with multi-variable inputs, one point can have an infinite amount of tangent lines. In two-variable maps, these lines form a tangent plane. More generally, a multi-variable map has a *tangent space* at each point. We can define this similarly for differentiable manifolds as the set of all tangent vectors at a certain point.

Definition 2.8. The set

$$T_pM := \{c^0 | c : (-\epsilon, \epsilon) \to M \text{ differentiable with } c(0) = p\}$$

is called the **tangent space** of M at the point p.

Before we start talking about the derivative of a map between differentiable manifolds, we prove the following lemma.

Lemma 2.9. Let M be an $n-$dimensional differentiable manifold, let $p \in M$ and let $x : U \to V$ be a chart of M with $p \in U$. Then the map

$$dx|_p : T_pM \to \mathbb{R}^n, \quad c^0 \mapsto \frac{d}{dt}(x \circ c)|_{t=0}$$

is well-defined and bijective.

Proof. It follows immediately from the equivalence relation that defines c^0 that the map is well-defined and injective. Let $v \in \mathbb{R}^n$ and set $c(t) := x^{-1}(x(p) + tv)$. Choose $\epsilon > 0$ such that $x(p) + tv \in V$ as long as $|t| < \epsilon$. We obtain

$$dx|_p(c^0) = \frac{d}{dt}\big(x \circ x^{-1}(x(p) + tv)\big)|_{t=0} = \frac{d}{dt}(x(p) + tv)|_{t=0} = v,$$

which shows surjectivity. □

Remark. We can equip T_pM with a vector space structure so that $dx|_p$ becomes a linear isomorphism. This structure is independent of choice of chart.

Now that we have established a linear approximation of M at p as T_pM it makes sense to define a derivative of a map f.

Definition 2.10. Let M and N be differentiable manifolds, let $p \in M$ and let $f : M \to N$ be differentiable near p. Then the derivative of f in p is given by the **differential of f in the point** p defined as the map

$$df|_p : T_pM \to T_{f(p)}N, \quad c^0 \mapsto (f \circ c)^0.$$

Remark. The differential of charts $x : U \to V$, i.e. $dx|_p$ now has two meanings: the map defined in Lemma 2.9 maps c^0 onto $\frac{d}{dt}(x \circ c)|_{t=0} \in \mathbb{R}^n$, while the differential of x in Definition 2.10 maps it onto $(x \circ c)^0 \in T_{x(p)}\mathbb{R}^n$. It is easy to see that the map $\mathbb{R}^n \to T_pM, \quad v \mapsto c^0_{p,v}$ with $c_{p,v}(t) := p + tv$ is an isomorphism relating the two definitions.

Lemma 2.11. $df|_p$ is well-defined and linear.

Proof. Let $x : U \to V$ be a chart of M such that $p \in U$ and let $y : \tilde{U} \to \tilde{V}$ be a chart of N such that $f(p) \in \tilde{U}$. Let $c^0 \in T_pM$. We show that $df|_p$ is independent of the choice $c : (-\epsilon, \epsilon) \to M$. By the chain rule, we calculate

$$
\begin{aligned}
dy|_{f(p)}((f \circ c)^0) &= (y \circ f \circ c)^0 \\
&= \big((y \circ f \circ x^{-1})^0 \circ (x \circ c)^0 \\
&= D(y \circ f \circ x^{-1})|_{x(p)} \cdot \big((x \circ c)^0\big) \\
&= D(y \circ f \circ x^{-1})|_{x(p)} \cdot dx|_p(c^0).
\end{aligned}
$$

Hence we obtain that

$$df|_p = \big(dy|_{f(p)}\big)^{-1} \circ D(y \circ f \circ x^{-1})|_{x(p)} \circ dx|_p$$

which is independent of the choice of representative c, hence it is well-defined. Moreover, it is the composition of three linear functions and therefore, the differential is linear. □

Theorem 2.12. ***(Chain Rule).*** *Let M, N and X be differential manifolds and let $p \in M$. Let $f : M \to N$ be differentiable near p and let $g : N \to X$ be differentiable near $f(p)$. Then,*

$$d(g \circ f)|_p = dg|_{f(p)} \circ df|_p.$$

Proof. Let $c : (-\epsilon, \epsilon) \to M$ with $c(0) = p$ be a differentiable curve. Then,

$$d(g \circ f)|_p(c^0) = \frac{d}{dt}\big((g \circ f) \circ c\big)|_{t=0} = \frac{d}{dt}\big(g \circ (f \circ c)\big)|_{t=0}$$

and we see that this is per definition equal to

$$dg|_{f(p)}\big((f \circ c)^0\big) = dg|_{f(p)}\big(df|_p(c^0)\big).$$

□

We see that the differential behaves like the derivative as we know it.

Theorem 2.13. *(**Inverse Function Theorem**). Let M and N be differentiable manifolds and $p \in M$. Let $f : M \to N$ be a C^k map $(k \geq 1)$. Assume that $df|_p : T_pM \to T_{f(p)}N$ is an isomorphism. Then there exists an open neighborhood $U \subset M$ of p and an open neighboorhood $W \subset N$ of $f(p)$ such that*

$$f|_U : U \to W$$

is a C^k-diffeomorphism.

Proof. Let $x : U_1 \to V_1$ be a chart of M with $p \in U_1$ and let $y : U_2 \to V_2$ be a chart of N with $f(p) \in U_2$. Then the map $y \circ f \circ x^{-1}$ is well-defined on $x\big(U_1 \cap f^{-1}(U_2)\big)$. Since $df|_p$ is an isomorphism, it is invertible, ensuring invertibility of total derivative $D(y \circ f \circ x^{-1})$. The "usual" inverse function theorem from multivariable analysis tells us that there exists open neighborhoods $V \subset x\big(U_1 \cap f^{-1}(U_2)\big)$ of $x(p)$ and $V' \subset V_2$ of $y\big(f(p)\big)$ so that $(y \circ f \circ x^{-1})|_V : V \to V'$ is a C^k-diffeomorphism. By setting $U := x^{-1}(V)$ and $W := y^{-1}(V')$ the assertion follows. □

Apart from partial derivatives, in multi-variable calculus, we can define a derivative of a function in an arbitrary direction. This can be extended to manifolds as well for arbitrary differentials and classes of the tangent space at a certain point.

Definition 2.14. Let M be a differential manifold, let $p \in M$ and let $c^0 \in T_pM$ corresponding to curve $c : (-\epsilon, \epsilon) \to M$. Let $f : M \to \mathbb{R}$ be differentiable near p. Then, the **directional derivative** of f in the direction c^0 is defined as

$$\partial_{c^0} f := df|_p(c^0) = \frac{d}{dt}(f \circ c)|_{t=0} \in \mathbb{R}.$$

2.3 Derivations

The goal of this section is to construct a useful manner to describe elements in the tangent space of a certain differentiable manifold M at point $p \in M$ in terms of charts x and the associated $dx|_p$, and how different charts can describe the same tangent space element. To do this, we introduce the notion of *derivations*, functions that behave similar to a derivative as we know it. First, let us cover a bit of notation.

Notation 2.15. Let $U \subset M$ be open and let $k \in \mathbb{N} \cup \{\infty\}$. We write

$$C^k(U) := \{f : U \to \mathbb{R} | f \text{ is } C^k\}.$$

Let $a \in \mathbb{R}$, let $f \in C^k(U)$ and let $g \in C^k(\tilde{U})$. Then we write

$$\begin{aligned} a \cdot f &\in C^k(U) & (a \cdot f)(z) &:= a \cdot f(z) \\ f + g &\in C^k(U \cap \tilde{U}) & (f + g)(z) &:= f(z) + g(z) \\ f \cdot g &\in C^k(U \cap \tilde{U}) & (f \cdot g)(z) &:= f(z) \cdot g(z) \end{aligned}$$

and we write

$$C^\infty_p := \bigcup_{U \text{ open},\ p \in U} C^\infty(U).$$

Definition 2.16. A map $\partial : C^\infty_p \to \mathbb{R}$ is called a **derivation** if the following holds:

(i) *Locality*: if $\tilde{U} \subset U$ is open and $p \in \tilde{U}$, then for $f \in C^\infty(U)$ we have

$$\partial f = \partial(f|_{\tilde{U}}).$$

In other words, zooming in further near p does not change the derivation of a function f.

(ii) *Linearity*: Let $a, b \in \mathbb{R}$ and let $f, g \in C^\infty_p$. Then we have

$$\partial(af + bg) = a\partial f + b\partial g.$$

(iii) *Leibniz Rule*: Let $f, g \in C_p^\infty$. Then,

$$\partial(f \cdot g) = \partial f \cdot g(p) + f(p) \cdot \partial g,$$

which can be compared to the product rule for differentiating.

To familiarize the reader with derivations, we give a few examples.

Example 2.17. (1) If $M = \mathbb{R}^n$, then $\frac{\partial}{\partial x^i}\big|_p$ is a derivation for all $1 \leq i \leq n$.

(2) Directional derivative ∂_{c^0} is a derivation. (i) and (ii) are obvious, since zooming in does not affect $c^0 \in T_pM$ and the differential $df|_p$ is linear. We check the Leibniz Rule (iii).

$$\begin{aligned}\partial_{c^0}(f \cdot g) &= \frac{d}{dt}\big((f \cdot g) \circ c\big)|_{t=0} \\ &= \frac{d}{dt}\big((f \circ c) \cdot (g \circ c)\big)|_{t=0} \\ &= \frac{d}{dt}(f \circ c)|_{t=0} \cdot g(c(0)) + \frac{d}{dt}(g \circ c)|_{t=0} \cdot f(c(0))\end{aligned}$$

where we used the product rule in the last equality. Using that $c(0) = p$ and the definition of the directional derivative, we see that this equals

$$\partial_{c^0}(f) \cdot g(p) + f(p) \cdot \partial_{c^0}(g).$$

Definition 2.18. The set $\mathbf{Der}(C_p^\infty)$ is the set of all derivations at $p \in M$.

Now, why would we want to look at the set of all derivations at a certain point? Let us get back to the directional derivative defined in Definition 2.14. Such a derivative, naturally, depends on the chosen direction $c^0 \in T_pM$. Therefore, we can view $\partial. := \partial_{\{.\}}$ as a function that takes elements of T_pM to the corresponding directional derivative. As we have seen in Example 2.17(ii), the directional derivative is indeed a derivation at p, hence an element of $\text{Der}(C_p^\infty)$. As will be shown shortly, this function is actually an isomorphism. This means that we can relate any element of the tangent space at p to a directional derivative at p. Before proving this, let us show that $\partial.$ is a linear map.

Lemma 2.19. The map $\partial. : T_pM \to \text{Der}(C_p^\infty), c^0 \mapsto \partial_{c^0}$ is linear.

Proof. Let $x : U \to V$ be a chart of M with $p \in U$ and let $f \in C^\infty(U)$. Let $v \in \mathbb{R}^n$ and define $c(t) := x^{-1}(x(p) + tv)$. Since $(dx|_p)^{-1}$ is linear, it suffices to show that $\partial. \cdot (dx|_p)^{-1}$ is. We have

$$\begin{aligned}\big(\partial. \circ (dx|_p)^{-1}(v)\big)(f) &= df|_p\big((dx|_p)^{-1}(v)\big) \\ &= df|_p\big(c^0\big) \\ &= \frac{d}{dt}\big(f \circ c\big)|_{t=0} \\ &= \frac{d}{dt}\big(f \circ x^{-1}(x(p) + tv)\big)|_{t=0} \\ &= \big\langle \text{grad}(f \circ x^{-1})|_{x(p)},\ v\big\rangle\end{aligned}$$

where grad denotes the usual gradient from multi-variable calculus and $\langle . \rangle$ denotes the Euclidean inner product. Since $\text{grad}(f \circ x^{-1})|_{x(p)}$ is independent of v, this expression is linear in v and hence this concludes the proof. ☐

Collorary 2.20. Let $\{e_i\}_{1\leq i\leq n}$ be the standard basis of $\mathbb{R}^n$ and let $x : U \to V$ be a chart of M such that $p \in U$. Then, $\{(dx|_p)^{-1}(e_i)\}_{1\leq i\leq n}$ is a basis for T_pM.

Proof. Let $f \in C^\infty(U)$ and let $c_i^0 := (dx|_p)^{-1}(e_i)$. We find

$$\partial_{c_i^0}(f) = \langle \text{grad}(f \circ x^{-1})|_{x(p)},\ e_i \rangle = \left.\frac{\partial(f\circ x^{-1})}{\partial x^i}\right|_{x(p)}.$$

□

Notation 2.21. We write, for charts $x : U \to V$ of M with $p \in U$ and maps $f \in C_p^\infty$

$$\left.\frac{\partial f}{\partial x^i}\right|_p := \left.\frac{\partial(f\circ x^{-1})}{\partial x^i}\right|_{x(p)}.$$

From Collorary 2.20 it becomes clear that, using the above notation, $\frac{\partial}{\partial x^i}|_p$ is a derivation for all $1 \leq i \leq n$.

Now let us prove that the tangent space at a point in a differentiable manifold is isomorphic to the set of derivations at that point.

Proposition 2.22. Let M be a differentiable manifold and let $p \in M$. Then, the map

$$\partial. : T_pM \to \text{Der}(C_p^\infty), \quad c^0 \mapsto \partial_{c^0}$$

is an isomorphism. More precisely, every derivation in $\text{Der}(C_p^\infty)$ is a directional derivative. Moreover, for every chart $x : U \to V$ with $p \in U$,

$$\left\{ \left.\frac{\partial}{\partial x^i}\right|_p \right\}_{1\leq i\leq n}$$

is a basis of $\text{Der}(C_p^\infty)$.

For this rather involved proof, we are going to need the following.

Lemma 2.23. Let $q \in \mathbb{R}^n$, let $r > 0$ and let $U := B(q, r)$. Now, let $h \in C^\infty(U)$. Then there exist functions $g_1, \dots, g_n \in C^\infty(U)$ such that

(i) $h(x) = h(q) + \sum_{i=1}^n (x^i - q^i) g_i(x)$, and

(ii) $\frac{\partial h}{\partial x^i}(q) = g_i(q)$.

Proof. (of Lemma 2.23). Let $x \in B(q, r)$. Define $c_x : [0, 1] \to \mathbb{R}$ by $c_x(t) := h(tx + (1 - t)q)$, i.e. the image under h of the straight line between q and x. We see that

$$\begin{aligned} h(x) - h(q) &= c_x(1) - c_x(0) \\ &= \int_0^1 \frac{d}{dt} c_x(\tau) d\tau \\ &= \int_0^1 \sum_{i=0}^n \frac{\partial h}{\partial x^i}\Big|_{tx+(1-t)q} (x^i - q^i) \\ &= \sum_{i=1}^n (x^i - q^i) \int_0^1 \frac{\partial h}{\partial x^i}\Big|_{tx+(1-t)q} dt \end{aligned}$$

where we used the chain rule in the third equality. Now, define

$$g_i(x) := \int_0^1 \frac{\partial h}{\partial x^i}\Big|_{tx+(1-t)q} dt.$$

This proves (i). Differentiating both sides of (i) at q w.r.t. x^i yields the expression in (ii). ☐

Proof. (of Proposition 2.22). It suffices to show that $\{\frac{\partial}{\partial x^i}\big|_p\}_{1\leq i\leq n}$ is a basis of $\mathrm{Der}(C_p^\infty)$. Indeed, by Collorary 2.20, the linear map $\partial.$ maps the basis of T_pM $\{(dx|_p)^{-1}(e_i)\}_i$ onto $\{\frac{\partial}{\partial x^i}\big|_p\}_i$. If this were the basis of $\mathrm{Der}(C_p^\infty)$, this linear map would send the basis of T_pM to the basis of $\mathrm{Der}(C_p^\infty)$, hence it is an isomorphism.
To prove this, we must show linear independence and the generating property (i.e. any element is a linear combination of the basis) of $\{\frac{\partial}{\partial x^i}\big|_p\}_{1\leq i\leq n}$.

(a) *Linear Independence.* Let $\gamma := \sum_{i=1}^n a^i \frac{\partial}{\partial x^i}\big|_p = 0$. It suffices to show that $a^i = 0$ for all $1 \leq i \leq n$. Let γ act on $f := x^j$. Then

$$0 = \sum_{i=1}^n a^i \frac{\partial x^j}{\partial x^i}\Big|_p = \sum_{i=1}^n a^i \delta_i^j = a^j, \quad \forall 1 \leq j \leq n$$

with Kronecker Delta δ_i^j.

(b) *Generating Property.* Let $\xi \in \mathrm{Der}(C_p^\infty)$. Define $a^j := \xi(x^j)$ for all $1 \leq j \leq n$. We are left to show that

$$\xi = \sum_{j=1}^n a^j \frac{\partial}{\partial x^j}.$$

First, we show that ξ vanishes on all constant functions. By Leibniz Rule we see that

$$\xi(1) = \xi(1 \cdot 1) = \xi(1) \cdot 1 + 1 \cdot \xi(1) = 2\xi(1)$$

hence $\xi(1) = 0$. Now, let $a \in \mathbb{R}$. We find, by the linearity property of derivations, that

$$\xi(a) = \xi(a \cdot 1) = a \cdot \xi(1) = 0.$$

Next, let $p \in U'$ and $U' \subset M$ open. Next, choose a neighborhood W of p such that $W \subset U \cap U'$ and $x(W) = B(x(p), r)$ for some $r > 0$. Now, let $f \in C^\infty(U') \subset C_p^\infty$. We apply Lemma 2.23 with $h := f \circ x^{-1}$. There must exist $g_1, \ldots, g_n \in C^\infty(B(x(p), r))$ such that

$$(f \circ x^{-1})(x) = (f \circ x^{-1})(x(p)) + \sum_{i=1}^n (x^i - x^i(p)) \cdot g_i(x), \quad \text{and}$$

$$\frac{\partial (f \circ x^{-1})}{\partial x^i}(x(p)) = g_i(x(p)).$$

We will now denote (i), (ii) and (iii) to be the properties of a derivation defined in Definition 2.16. It follows now that

$$\begin{aligned}\xi(f) &\overset{\text{(i)}}{=} \xi(f|_W)\\ &= \xi\Big(f(p) + \sum_{i=1}^{n} \big(x^i - x^i(p)\big)(g_i \circ x)\Big)\\ &\overset{\text{(ii)}}{=} \sum_{i=1}^{n} \xi\Big(\big(x^i - x^i(p)\big)(g_i \circ x)\Big) \quad \text{since } \xi(f(p)) \text{ vanishes}\\ &\overset{\text{(iii)}}{=} \sum_{i=1}^{n} \Big(\xi\big(x^i - x^i(p)\big) \cdot g_i(x(p)) + \big(x^i - x^i(p)\big)|_p \cdot \xi(g_i \circ x)\Big).\end{aligned}$$

Now we see $\xi(x^i(p)) = 0$ as $x^i(p)$ is constant. Also, evaluating $(x^i - x^i(p))$ in p yields 0. Hence, by linearity of ξ (property (ii)), we see that

$$\begin{aligned}\xi(f) &= \sum_{i=1}^{n} \xi(x^i) g_i(x(p))\\ &= \sum_{i=1}^{n} a^i \cdot \frac{\partial f}{\partial x^i}\bigg|_p\end{aligned}$$

which points out that the generating property is satisfied. Hence, we have completed the proof.

□

Hence, we have concluded from Lemma 2.9 and Proposition 2.22 that for any differentiable manifold M and $p \in M$,

$$\begin{aligned}\mathbb{R}^n \overset{(dx|_p)^{-1}}{\cong} T_pM \overset{\partial_\cdot}{\cong} \text{Der}(C_p^\infty)\\ e_i \mapsto (dx|_p)^{-1}(e_i) \mapsto \frac{\partial}{\partial x^i}\bigg|_p.\end{aligned}$$

From now on, we will identify tangent vectors with their corresponding directional derivative as follows.

Notation 2.24. Let $\xi \in T_pM$. We write

$$\xi := \partial_\xi = \sum_{i=1}^{n} \xi^i \frac{\partial}{\partial x^i}\bigg|_p$$

with $\xi^i = (dx|_p(\xi))^i$, instead of ∂_ξ or $\xi = \sum_{i=1}^{n} \xi^i (dx|_p)^{-1}(e_i)$.

Now, we can ask ourselves: how do the coefficients $\xi^i, \dots, \xi^n$ of a tangent vector change under transformation of charts? Let $y : \tilde{U} \to \tilde{V}$ be another chart. We express the same tangent vector $\xi \in T_pM$ with respect to both charts, i.e.

$$\xi = \sum_{i=1}^{n} \xi^i \frac{\partial}{\partial x^i}\bigg|_p = \sum_{i=1}^{n} \eta^i \frac{\partial}{\partial y^i}\bigg|_p$$

and we are curious how the coefficients ξ^i depend on coefficients η^i and vice versa. First, we observe that

$$(\xi^1, \dots, \xi^n)^T = dx|_p(\xi) = (dx|_p)\Big((dy|_p)^{-1}\big((\eta^1, \dots, \eta^n)^T\big)\Big).$$

Now, we apply the Chain Rule from Theorem 2.12 to compute

$$(\xi^1, \dots, \xi^n)^T = D(x \circ y^{-1})|_{y(p)}(\eta^1, \dots, \eta^n)^T$$

in which we can swap the roles of x and y to obtain

$$(\eta^1, \dots, \eta^n)^T = D(y \circ x^{-1})|_{x(p)}(\xi^1, \dots, \xi^n)^T.$$

Hence, we obtain

$$\eta^j = \sum_{i=1}^{n} \frac{\partial(y^j \circ x^{-1})}{\partial x^i}\bigg|_{x(p)} \xi^i. \tag{2.2}$$

Let us briefly interpret this result from a physicist's point of view. Often, tangent vectors are referred to as *contravariant vectors*. The word contravariant indicates that, under a change of basis, the vector changes in the exact "opposite" way to compensate for the change. For example, if a basis of $\mathbb{R}^n$ is changed under a certain matrix, then the contravariant vector changes under its inverse. A change of basis is essentially a change of coordinate system. This has many applications in physics, for example when studying mechanics (or any other field of classical/special relativistic physics) in different reference frames.
In differential geometry, charts can essentially be viewed as a coordinate system relating a neighborhood of a point in a manifold to the Euclidean space. Equation (2.2) tells us how the coefficients of a certain tangent vector in the tangent space are transformed as we choose another chart, or in other words, a different coordinate system. In many fields of physics (of which general relativity is the most famous example), space (or spacetime) is often described by manifolds (curved spacetime). Hence, this transformation rule is of great importance for describing physical phenomena such as gravity. To illustrate how the transformation works, we can look at a special case.

Example 2.25. Let $\xi = \frac{\partial}{\partial x^i}\big|_p$, or equivalently $(\xi^1, \dots, \xi^n)^T = e_i$. We want to know how to describe ξ using another chart y. By applying Equation (2.2), we find

$$\begin{aligned}\frac{\partial}{\partial x^i}\bigg|_p &= \sum_{j=1}^{n} \eta^j \frac{\partial}{\partial y^j}\bigg|_p \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} \xi^k \frac{\partial(y^j \circ x^{-1})}{\partial x^k}\bigg|_{x(p)} \cdot \frac{\partial}{\partial y^j}\bigg|_p.\end{aligned}$$

Now, we see $\xi^k = \delta_i^k$ and hence the summation over k is cancelled and we can replace $k \to i$. We obtain

$$\frac{\partial}{\partial x^i}\bigg|_p = \sum_{j=1}^{n} \frac{\partial(y^j \circ x^{-1})}{\partial x^i}\bigg|_{x(p)} \cdot \frac{\partial}{\partial y^j}\bigg|_p. \tag{2.3}$$

Equation (2.3) essentially tells us how a basis of directional derivatives changes under a change of charts.

Describing tangent vectors with respect to the basis $\frac{\partial}{\partial x^i}\big|_p$ requires writing out summations over certain indices, as can be seen in the discussion above. In physics literature, it is common to use the *Einstein summation convention* to shorten expressions. The convention means that when an index appears twice in an expression (necessarily once as an upper index and once as a lower index), then a summation over this index is implicit. In this way, Equation (2.3) is written as

$$\frac{\partial}{\partial x^i}\bigg|_p = \frac{\partial(y^j \circ x^{-1})}{\partial x^i}\bigg|_{x(p)} \cdot \frac{\partial}{\partial y^j}\bigg|_p.$$

Note here that the index j in $\frac{\partial}{\partial y^j}$ is considered a lower index and can be interpreted as "the derivative with respect to y_j". Often, this expression is written even shorter using Notation 2.21 and leaving out the indication of the point p, as

$$\frac{\partial}{\partial x^i} = \frac{\partial y^j}{\partial x^i} \cdot \frac{\partial}{\partial y^j}.$$

This expression is relatively easy to remember, since we can cancel ∂y^j. The Einstein summation convention will be used throughout this thesis.

2.4 Vector Fields

We have established a formalism to describe derivatives of functions between differentiable manifolds in a certain point $p \in M$. However, in some cases we want to describe behaviour of functions at all points of a manifold at once. In vector spaces, functions that associate a direction and magnitude (i.e. a vector) to all points in the space are called *vector fields*. For differentiable manifolds, it is a bit more subtle, since each point has its own distinct tangent space. Hence, we define a vector field as a function that assigns to each point in a differentiable manifold a tangent vector of the associated tangent space. To define it properly, we are going to need the set of all tangent spaces.

Definition 2.26. Let M be a differentiable manifold. The **Tangent bundle** of M is defined as

$$TM := \bigcup_{p \in M} T_pM.$$

As it turns out, we can equip TM with a differentiable manifold structure. Let $\pi : TM \to M$ be the "projection" map that sends a tangent vector belonging to a tangent space T_pM for a $p \in M$ to the point p itself. Let A be the differentiable structure of M. For every chart $x : U \to V$ in A we can construct a chart $X_x : U_x \to V_x$ of TM, where

$$\begin{aligned} U_x &:= \pi^{-1}(U) \subset TM, \\ V_x &:= V \times \mathbb{R}^n \subset \mathbb{R}^{2n} \quad \text{and} \\ X_x(\xi) &:= \Big(x\big(\pi(\xi)\big), dx|_{\pi(\xi)}(\xi)\Big) \end{aligned}$$

so that we have $X_x^{-1}(v, w) = (dx|_{x^{-1}(v)})^{-1}(w)$. We will not go into the proof that TM is indeed a differentiable manifold. For the proof, we refer to [2]. However, this fact does allow us to define the notion of a vector field in differential geometry. As a final remark, we can express the projection map $\pi : TM \to M$ in terms of charts x as

$$x \circ \pi \circ X_x^{-1} : V \times \mathbb{R}^n \to V, \ \ (v, w) \mapsto v$$

from which we see that it is a smooth map.

Definition 2.27. A map $\eta : M \to TM$ is a **vector field** on M if

$$\pi\big(\eta(p)\big) = p$$

for all $p \in M$. Vector fields are sometimes also called *sections of the tangent bundle*.

This definition says that such a function η is a vector field if it assigns to each p a tangent vector from the *correct* tangent space T_pM, or equivalently, if the tangent vector "projects" to p under π. In addition, since both M and TM are differentiable manifolds, we already know what it means for a vector field to be differentiable, or C^k.
We can characterize a vector field by coefficient functions, rather than coefficients. Let $x : U \to V$ be a chart of M. Then, a vector field η on U has coefficient functions $\eta^1, \dots, \eta^n : V \to \mathbb{R}$ so that

$$\eta(p) = \xi^i\big(x(p)\big) \frac{\partial}{\partial x^i}\bigg|_p$$

(notice the Einstein summation convention). We say that η is C^k if and only if its coefficient functions are.
Let us look at an example of a vector field.

Example 2.28. Consider $M = \mathbb{R}^2$ with polar coordinates. Let $\phi_0 \in \mathbb{R}$. Set $U := \mathbb{R}^2 \backslash l$ where l denotes the half-line from the origin under angle ϕ_0 from the horizontal axis, i.e. $l := \{\lambda \cdot (\cos \phi_0, \sin \phi_0)^T | \lambda \in \mathbb{R}_{\geq 0}\}$. Next, set $V := (0, \infty) \times (\phi_0, \phi_0 + 2\pi)$. Define the function $F : V \to U$, $(r, \phi) \mapsto (r \cos \phi, r \sin \phi)$ so that we can choose chart $y := F^{-1}$.
On U we can define the vector field $\eta := r \frac{\partial}{\partial r}$, such that its coefficient functions are

$$\eta^1(r, \phi) = r$$
$$\eta^2(r, \phi) = 0.$$

Tangent vectors associated to points in M are pointed radially outward and its magnitude is equal to the distance from the origin r. This can be seen in Figure 2.1(a). Since we chose our manifold to be a Euclidean space, an obvious choice for a chart is the identity function $x := \text{id} : \mathbb{R}^2 \to \mathbb{R}^2$. We wish to express η in terms of x instead of y, where we express x^1 and x^2 as $r \cos \phi$ and $r \sin \phi$, respectively. To do this, we use Equation (2.3) to find

$$\begin{aligned}
\eta &= r \frac{\partial}{\partial r} \\
&= r \left(\frac{\partial x^1}{\partial r} \frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial r} \frac{\partial}{\partial x^2} \right) \\
&= r \left(\cos \phi \frac{\partial}{\partial x^1} + \sin \phi \frac{\partial}{\partial x^2} \right) \\
&= x^1 \frac{\partial}{\partial x^1} + x^2 \frac{\partial}{\partial x^2}.
\end{aligned}$$

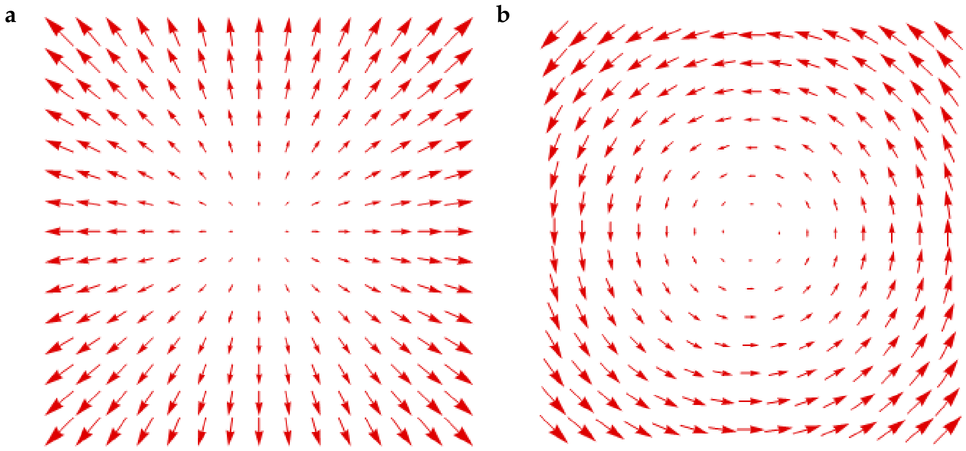

Figure 2.1: Vector Fields in the Euclidean plane. (**a**) The vector field $\eta = r\frac{\partial}{\partial r}$. (**b**) The vector field $\eta = \frac{\partial}{\partial \phi}$.

Hence, in Cartesian coordinates we find coefficient functions

$$\xi^1(x^1, x^2) = x^1$$
$$\xi^2(x^1, x^2) = x^2.$$

Similarly, define a vector field on U by $\eta := \frac{\partial}{\partial \phi}$. This vector field assigns a vector to each point that points in the axial direction (perpendicular to the radial direction; in the direction of the "angular frequency") and its magnitude is proportional to the distance to the origin, as shown in Figure 2.1(b). Here, the coefficient functions in polar coordinates are given by

$$\eta^1(r, \phi) = 0$$
$$\eta^2(r, \phi) = 1$$

and we can use Equation (2.3) to find the Cartesian coefficient functions. We compute

$$\begin{aligned}\frac{\partial}{\partial \phi} &= \frac{\partial x^1}{\partial \phi}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial \phi}\frac{\partial}{\partial x^2} \\ &= -r\sin\phi\frac{\partial}{\partial x^1} + r\cos\phi\frac{\partial}{\partial x^2} \\ &= -x^2\frac{\partial}{\partial x^1} + x^1\frac{\partial}{\partial x^2}.\end{aligned}$$

Hence, we find Cartesian coefficient functions

$$\xi^1(x^1, x^2) = -x^2$$
$$\xi^2(x^1, x^2) = x^1.$$

This concludes the chapter on differential geometry theory. The next chapter includes the theory on Riemannian Geometry, where notions of distance and angles between points on a differentiable manifold are introduced.

3. Riemannian Geometry

So far, we have defined differentiable functions on manifolds, giving rise to differentiable manifolds. This allows us to do analysis with linear approximations to manifolds (tangent spaces) and to maps (differentials). In order to describe the behaviour of solutions to the heat equation and the behaviour of Brownian motion on manifolds, we are required to develop a formalism that incorporates distance between points and angles between tangent vectors. In other words, we need to do geometry. To be able to do geometry on manifolds, we introduce scalar products on tangent spaces. The formalism that incorporates all of the above is Riemannian geometry and gives rise to Riemannian manifolds. This will be introduced in this chapter.

In Euclidean spaces, we are comfortable with the fact that heat from a point source travels radially outward from that point. For instance, think about the heat of a candle: if you hold your hand anywhere, but at a constant distance from the candle, you will feel the same amount of heat. The intuitive fact that heat doesn't travel along any arbitrary path, but in a straight path radially outward, follows from the fact that the straight path is the path along which the energy it takes to travel that distance is minimized (the Lagrangian action is minimized). One can imagine that on certain manifolds, for instance on a sphere, this energy-minimizing path between two points is not a straight path anymore. In general, an energy-minimizing path between two points on a manifold is called a *geodesic* (this is not the definition, but a property). Geodesics are important for studying the heat equation and Brownian motion and will be elaborated on as well in this chapter. The chapter will follow the structure of [2] and includes insights from [4].

3.1 Bilinear forms

Before diving into Riemannian geometry, we are going to need some basics from linear algebra. First we introduce the notion of a bilinear form.

Definition 3.1. Let V be an n-dimensional $\mathbb{R}$-vector space. A **symmetric bilinear form** is a map $g : V \times V \to \mathbb{R}$ that satisfies

(i) $g(av + bw, u) = ag(v, u) + bg(w, u)$ for all $v, w, u \in V$ and $a, b \in \mathbb{R}$ and

(ii) $g(v, w) = g(w, v)$ for all $v, w \in V$.

Hence, it is a symmetric function and linear in both the first and second parameter (due to symmetry). We call g **non-degenerate** if $g(v, w) = 0$ for all $w \in V$ implies that $v = 0$.

Let $(b_1, \ldots, b_n)$ be a basis of V. Then, we define

$$g_{ij} := g(b_i, b_j)$$

for all $i, j = 1, \ldots, n$. We see that the g_{ij} form a symmetric $n \times n$-matrix. Conversely, we can construct g from its matrix elements g_{ij}. Indeed, if $v = a^i b_i$ and $w = c^j b_j$ in

V (note the Einstein summation convention), we find

$$g(v,w) = g(a^i b_i, c^j b_j) \overset{\text{(ii)}}{=} a^i c^j g_{ij}.$$

Definition 3.2. Let V be an n-dimensional $\mathbb{R}$-vector space. Then the **dual space** is defined as

$$V^* := \{\text{linear maps } V \to \mathbb{R}\}$$

and the **dual basis** $(b_1^*, \ldots, b_n^*)$ of basis of V $(b_1, \ldots, b_n)$ satisfies so that $b_i^*(b_j) = \delta_{ij}$.

Notation 3.3. Often we write

$$g = \sum_{i,j=1}^{n} g_{ij} b_i^* \otimes b_j^*$$

so that for $v, w \in V$ we can write

$$g(v,w) = \sum_{i,j=1}^{n} g_{ij} b_i^*(v) \cdot b_j^*(w) = g_{ij} a^i c^j$$

as above.

Since $(g_{ij})_{i,j=1,\ldots,n}$ denotes a symmetric matrix with real elements, it is diagonalizable and hence there exists a basis $(e_1, \ldots, e_n)$ of V such that $g(e_i, e_j)$ is always zero if $i \neq j$ and is either -1 or 1 if $i = j$. Hence, we have

$$(g_{ij})_{i,j=1,\ldots,n} = \begin{pmatrix} -1 & & & & & 0 \\ & \ddots & & & & \\ & & -1 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ 0 & & & & & 1 \end{pmatrix}. \tag{3.1}$$

We call this basis the **generalized orthonormal basis**. Moreover, the **index** of g is defined as the amount of times -1 occurs on the diagonal in Equation (4.5), and is denoted $\text{Index}(g)$. Observe that g is positive definite if and only if $\text{Index}(g) = 0$. In that case, the expression in Equation (4.5) is the identity matrix. Therefore, the positive definiteness of g is equivalent with g being a Euclidean scalar product (w.r.t. the generalized orthonormal basis).

Let $B = (b_1, \ldots, b_n)$ be a basis of V. We can ask ourselves how the elements g_{ij} of the representing matrix of g change under a transformation of bases. Let $A = (a_1, \ldots, a_n)$ be another basis of V and denote elements of the transformation matrix T by t_i^j so that we have $a_i = t_i^j b_j$. Then,

$$\begin{aligned} g_{ij}^A &= g(a_i, a_j) \\ &= g(t_i^k b_k, t_j^l b_l) \\ &\overset{\text{(ii)}}{=} t_i^k t_j^l g(b_k, b_l) \\ &= t_i^k t_j^l g_{kl}^B. \end{aligned} \tag{3.2}$$

Along with bilinear forms come a kind of isomorphism between them.

Definition 3.4. Let V and W be n-dimensional $\mathbb{R}$-vector spaces and let g_V and g_W be symmetric bilinear forms on them. Then a map $\Phi : W \to V$ is called an **isometry** if it is linear, bijective and satisfies

$$g_V\big(\Phi(w_1), \Phi(w_2)\big) = g_W(w_1, w_2) \ \ \forall w_1, w_2 \in W,$$

that is, if $\Phi^* g_V = g_W$, where $(\Phi^* g)(w_1, w_2) := g\big(\Phi(w_1), \Phi(w_2)\big)$.

3.2 Riemannian metrics

We are now fully equipped to introduce the notion of Riemannian geometry.

Definition 3.5. Let M be a differentiable manifold. Let g be a map that assigns to each point $p \in M$ a non-degenerate symmetric bilinear form $g|_p$ on the associated tangent space T_pM. Let $x : U \to V$ be a chart of M. We define the function $g_{ij}^{(x)} := g_{ij} : V \to \mathbb{R}$ as

$$g_{ij}(v) := g|_{x^{-1}(v)}\left(\frac{\partial}{\partial x^i}\bigg|_{x^{-1}(v)}, \frac{\partial}{\partial x^j}\bigg|_{x^{-1}(v)}\right).$$

If g "depends smoothly on the base point", i.e. for every chart $x : U \to V$ of M the $g_{ij} : V \to \mathbb{R}$ are C^∞-functions, then g is called a **semi-Riemannian metric**. If, in addition, $g|_p$ is positive definite for all $p \in M$, then g is called a **Riemannian metric**.

We see that Riemannian metrics are a Euclidean scalar product by its positive definiteness.

Remark. Note that g can be viewed as a tensor field (a function that assigns a tensor, in this case a matrix, to all points on the manifold). Similar to vector fields, g is expressed with respect to the basis $\{\frac{\partial}{\partial x^i}\}_i$ of the tangent space (or actually of the set of derivations, see Notation 2.24) induced by the chart x. Moreover, similar to the smoothness of vector fields, we then require the coefficient functions to be smooth.

Definition 3.6. A pair (M, g) with differentiable manifold M and (semi-)Riemannian metric g is called a **(semi-)Riemannian manifold.**

3.2.1 Choice of charts

It is imperative to determine how a (semi-)Riemannian metric g transforms under change of charts. Let $x : U \to V$ and $y : \tilde{U} \to \tilde{V}$ be two distinct charts of M such that $p \in U \cap \tilde{U}$. We apply Equation (2.3) to find (using Einstein's summation convention)

$$\frac{\partial}{\partial y^i}\bigg|_p = \frac{\partial(x^j \circ y^{-1})}{\partial y^i}\bigg|_{y(p)} \cdot \frac{\partial}{\partial x^j}\bigg|_p.$$

Next, define $a_i := \frac{\partial}{\partial y^i}\big|_p$, $t_i^j := \frac{\partial(x^j \circ y^{-1})}{\partial y^i}\big|_{y(p)}$ and $b_i := \frac{\partial}{\partial x^j}\big|_p$. We apply Equation (3.2) to find

$$g_{ij}^{(y)}\big(y(p)\big) = \frac{\partial(x^k \circ y^{-1})}{\partial y^i}\bigg|_{y(p)} \cdot \frac{\partial(x^l \circ y^{-1})}{\partial y^j}\bigg|_{y(p)} \cdot g_{kl}^{(x)}\big(x(p)\big)$$

hence we must have for all $v \in y(U \cap \tilde{U})$

$$g_{ij}^{(y)}(v) = \frac{\partial(x^k \circ y^{-1})}{\partial y^i}\bigg|_v \cdot \frac{\partial(x^l \circ y^{-1})}{\partial y^j}\bigg|_v \cdot g_{kl}^{(x)}\big((x \circ y^{-1})(v)\big) \tag{3.3}$$

or, using Notation 2.21 and not noting the point in which the partial derivatives are defined (physicist's notation),

$$g_{ij}^{(y)} = \frac{\partial x^k}{\partial y^i}\frac{\partial x^l}{\partial y^j}\left(g_{kl}^{(x)} \circ (x \circ y^{-1})\right).$$

Recall that $dx|_p : T_pM \to \mathbb{R}^n$ is a linear isomorphism for all charts $x : U \to V$ with $p \in U$. Thus, so are all its components $dx^i|_p : T_pM \to \mathbb{R}, 1 \leq i \leq n$. By Definition 3.2, we see that $dx^i|_p \in (T_pM)^*$ for all $1 \leq i \leq n$.

Definition 3.7. The dual space $T_p^*M := (T_pM)^*$ is called the **cotangent space** of M at p.

Lemma 3.8. The dual basis of $\{\frac{\partial}{\partial x^i}\big|_p\}_i$ is $\{dx^i|_p\}_i$.

Proof. As we have seen in Proposition 2.16 we have that $dx|_p(\frac{\partial}{\partial x^i}\big|_p) = e_i$ and hence we have $dx^j|_p(\frac{\partial}{\partial x^i}\big|_p) = \delta_i^j$ for all $i, j = 1, \ldots, n$. □

This allows for a useful way to write g. Namely, according to Notation 3.3 we write

$$g|_p = g_{ij}\big((x(p)\big) \cdot dx^i|_p \otimes dx^j|_p$$

or in physicist's notation: $g|_p = g_{ij} \cdot dx^i \cdot dx^j$.

Lemma 3.8 allows us to express coefficients of $dx|_p$ of a chart x with respect to coefficients of $dy|_p$ of chart y (at this point, it goes without saying that p is assumed to be an element of the intersection of the domains of x and y).

Lemma 3.9. Let $(a_1, \ldots a_n)$ and $(b_1, \ldots, b_n)$ be bases of a vector space. Let t_i^j be elements of the transformation matrix T so that $a_i = t_i^j b_j$. Then the associated dual bases satisfy

$$b_i^* = t_i^j a_j^*.$$

Proof. We have

$$\begin{aligned} t_i^j a_j^*(b_k) &= \left(t_i^j a_j^*\right)\left((T^{-1})_k^l a_l\right) \\ &= t_i^j (T^{-1})_k^l a_j^*(a_l). \end{aligned}$$

Note that $a_j^*(a_l) = \delta_{jl}$ hence this expression is equal to $t_j^i (T^{-1})_k^j$ which essentially boils down to the element of the i-th column and the k-th row of the matrix $T(T^{-1})$, i.e. the identity matrix. Of course, this is 1 if the element is on the diagonal and 0 otherwise, hence this is equal to δ_k^i. Thus, $t_i^j a_j^*$ is by definition the i-th element of the dual basis of $(b_1, \ldots, b_n)$ and we must have that it equals b_i^*. □

Collorary 3.10. Let $b_i^* = dx^i|_p$ and $a_j^* = dy^j|_p$ for all $i, j = 1, \ldots, n$. Then we have

$$dx^i|_p = \frac{\partial(x^i \circ y^{-1})}{\partial y^j} \cdot dy^j|_p. \tag{3.4}$$

or in physicist's short notation: $dx^i = \frac{\partial x^i}{\partial y^j} dy^j$.

3.2.2 Examples of Riemannian metrics

The definition of the Riemannian metric is rather abstract. To familiarize the reader with Riemannian metrics, we discuss some examples in this section.

Example 3.11. The first intuitive example of a Riemannian metric is the Euclidean space $\mathbb{R}^n$ with standard basis $\{e_i\}_i := \{\frac{\partial}{\partial x^i}\}_i$ for chart $x = \text{id} : \mathbb{R}^n \to \mathbb{R}^n$ (we leave out the subscript p since this is the same at all points in $\mathbb{R}^n$). The Euclidean inner product says, for $v, w \in \mathbb{R}^n$ that $\langle v, w\rangle = v^j w_j$. For the metric g_{Eucl} we have

$$g_{\text{Eucl}}(v, w) = (g_{\text{Eucl}})_{ij} v^i w^j$$

thus we must have that $(g_{\text{Eucl}})_{ij}$ is the identity matrix.

But what if we chose polar coordinates? Let us take $n = 2$. Set dx^1 and dx^2 the components of the differential associated to chart $x = \text{id}$ (again omitting the subscript p). Then, according to Notation 3.3 we can write $g_{\text{Eucl}} = dx^1 \otimes dx^2$. Now, writing $x^1 = r\cos\phi$ and $x^2 = r\sin\phi$, we wish to express g_{Eucl} in terms of differential components dr and $d\phi$. By Equation (3.4), we see that

$$\begin{aligned} dx^1 &= \frac{\partial x^1}{\partial r} dr + \frac{\partial x^1}{\partial \phi} d\phi = \cos\phi\, dr - r\sin\phi\, d\phi \\ dx^2 &= \frac{\partial x^2}{\partial r} dr + \frac{\partial x^2}{\partial \phi} d\phi = \sin\phi\, dr + r\cos\phi\, d\phi. \end{aligned}$$

As a result, we find

$$\begin{aligned} g_{\text{Eucl}} &= (\cos\phi\, dr - r\sin\phi\, d\phi) \otimes (\cos\phi\, dr - r\sin\phi\, d\phi) \\ &\quad + (\sin\phi\, dr + r\cos\phi\, d\phi) \otimes (\sin\phi\, dr + r\cos\phi\, d\phi) \\ &= dr \otimes dr + r^2(d\phi \otimes d\phi). \end{aligned}$$

and hence we have

$$(g_{\text{Eucl}})_{ij}^{\text{polar}} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.$$

What this matrix says is that in any point $(r, r\sin\phi) \in \mathbb{R}^2$ we have that

- the tangent vector $\frac{\partial}{\partial r}$ has length 1,
- the tangent vector $\frac{\partial}{\partial \phi}$ has length r, and
- $\frac{\partial}{\partial r}$ and $\frac{\partial}{\partial \phi}$ are orthogonal to each other (since the off-diagonal elements are zero, indicating their Euclidean inner product is zero).

These results are exactly what we found in Example 2.28.

Another example is what happens when the Euclidean plane $\mathbb{R}^2$ is "bent" to a certain shape (for example, a paraboloid). To elaborate on this, we can determine in general what metrics on submanifolds look like.

Remember that for any n-dimensional differentiable manifold M we have, for each $p \in M$, a canonical isomorphism $\Phi_p : T_pM \to \mathbb{R}^n$ exists, defined by $c^0 \to \frac{d}{dt}c|_{t=0}$, where c^0 is the equivalence class of the curve $c : (-\epsilon, \epsilon)$. Thus, for an n-dimensional submanifold $M \subset \mathbb{R}^{n+k}$ with $k = 1, 2, \ldots$, $\Phi_p : T_pM \to \mathbb{R}^{n+k}$ becomes an injective canonical map.

Lemma 3.12. Let $M \subset \mathbb{R}^{n+k}$ be an n-dimensional submanifold with canonical injective map Φ_p. Then (M, g) is a Riemannian manifold, where the metric g is defined such that each $g|_p := \Phi_p^*\langle \cdot, \cdot \rangle$ (notation according to Definition 3.4), where $\langle x, y\rangle = \sum_{i=1}^{n+k} x^i y^i$ denotes the usual Euclidean scalar product for $x, y \in M$ (since they are points of $\mathbb{R}^{n+k}$ as well).

Proof. First of all, the Euclidean scalar product is positive definite. Since Φ_p is injective, we must have that $g|_p$ is positive definite as well for all $p \in M$ (indeed, $\Phi_p^*\langle x, x\rangle = \langle \Phi_p(x), \Phi_p(x)\rangle \geq 0$ and by injectivity we have $\Phi_p^*\langle x, x\rangle = 0 \implies \Phi_p(x) = 0$, hence x must be 0, showing positive definiteness). Hence, we see that (M, g) with the Riemannian metric defined using $g|_p$ as in Definition 3.5 is a Riemannian manifold for all submanifolds M. □

Definition 3.13. This metric g is called the **first fundamental form**.

Lemma 3.14. The g_{ij} of the first fundamental form are smooth.

Proof. The charts of submanifolds correspond to local parametrizations of M, which means maps $F : V \to M$ with $V \subset \mathbb{R}^n \subset \mathbb{R}^{n+k}$ open, so that $x = F^{-1} : U = F(V) \to V$ is a chart of M. Now, let $p = x^{-1}(v)$ for some $v \in V$. Then we have

$$\begin{aligned} g_{ij}(v) &= g|_p\left(\frac{\partial}{\partial x^i}\bigg|_p, \frac{\partial}{\partial x^j}\bigg|_p\right) \\ &= \left(\Phi_p^*\langle \cdot, \cdot\rangle\right)\left(\frac{\partial}{\partial x^i}\bigg|_p, \frac{\partial}{\partial x^j}\bigg|_p\right) \\ &= \left\langle \Phi_p\left(\frac{\partial}{\partial x^i}\bigg|_p\right), \Phi_p\left(\frac{\partial}{\partial x^j}\bigg|_p\right)\right\rangle. \end{aligned}$$

Now, we remember that the tangent vector $\frac{\partial}{\partial x^i}\big|_p$ is essentially the equivalence class of a curve through p that is the preimage under the chart x of a curve through $x(p) = v$ defined by $c(t) = v + te_i$ (with standard Euclidean basis $(e_1, \dots e_n)$). This preimage is then exactly $F \circ c$, thus we see that $\Phi_p\big(\frac{\partial}{\partial x^i}\big|_p\big) = \frac{d}{dt}F(v+te_i)|_{t=0}$, which is by definition $\frac{\partial F}{\partial x^i}(v)$ for all $1 \leq i \leq n$. As a result, we have

$$g_{ij}(v) = \langle \frac{\partial F}{\partial x^i}(v), \frac{\partial F}{\partial x^j}(v)\rangle$$

and we obtain

$$g_{ij} = \langle \frac{\partial F}{\partial x^i}, \frac{\partial F}{\partial x^j}\rangle \tag{3.5}$$

which is a smooth function. □

To clarify this discussion, let us look at an example of a submanifold of $\mathbb{R}^3$.

Example 3.15. Let $P := \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 = 2z\} \subset \mathbb{R}^3$ be the paraboloid as shown in Figure 3.1. We can obtain P by "bending" the Euclidean plane $\mathbb{R}^2$ to the proper shape. This "bending" can be seen as a parametrization from $\mathbb{R}^2$ to P, given by

$$\begin{aligned} F : V = \mathbb{R}^2 &\to P \\ (x, y) &\mapsto \left(x, y, \frac{1}{2}(x^2 + y^2)\right). \end{aligned}$$

As is illustrated in Figure 3.1, the basis of P given by $\left(\frac{\partial}{\partial x^1}\big|_p, \frac{\partial}{\partial x^2}\big|_p\right) =: \left(F(e_1)|_p, F(e_2)|_p\right)$ for $p \in P$ induced by chart $x = F^{-1}$ (the orthogonal projection onto the plane $z = 0$) is "stretched out" in points $p = F((x,y))$ with $(x,y) \in \mathbb{R}^2$ further away from the origin, and they need not be orthogonal with respect to the Euclidean scalar product on Cartesian coordinates. To verify this, let us compute g_P. By Equation (3.5), we see that, for $p = F((x,y))$,

$$(g_P)_{ij}|_p = \begin{pmatrix} 1+x^2 & xy \\ xy & 1+y^2 \end{pmatrix}.$$

This tells us the following:

- if (x,y) is further away from the origin, the distance between points grows larger,
- since the off-diagonal terms are non-zero everywhere except in the origin, the basis $\left(F(e_i)|_p, F(e_j)|_p\right)$ is not orthogonal,
- close to the origin/centre of P (i.e. in the limit of small x, y) g_P resembles the Euclidean metric. In Figure 3.1 this can be seen as the paraboloid being "tangent" to the Euclidean plane, meaning that if you zoom in far enough, the "bending" is negligible.

We can also determine g_P in terms of cylindrical coordinates. In this case, we can define $P = \{(r\cos\phi, r\sin\phi, z) \in \mathbb{R}^3 | r^2 = 2z\}$. By Equation (3.4) we see that $dz = r\,dr$. Since cylindrical coordinates are essentially polar coordinates extended to $\mathbb{R}^3$ by Cartesian coordinate z, we can use the calculations from Example 3.11 to find

$$\begin{aligned} g_P &= dr \otimes dr + r^2(d\phi \otimes d\phi) + dz \otimes dz \\ &= dr \otimes dr + r^2(d\phi \otimes d\phi) + r^2(dr \otimes dr) \\ &= (1+r^2)(dr \otimes dr) + r^2(d\phi \otimes d\phi) \end{aligned}$$

and hence we find

$$g_P^{\text{cyl}} = \begin{pmatrix} 1+r^2 & 0 \\ 0 & r^2 \end{pmatrix}.$$

We see that $\frac{\partial}{\partial r}$ and $\frac{\partial}{\partial \phi}$ are orthogonal and that the metric grows larger for points further away from the origin. Note that, in this case, the metric gives problems in the origin of $\mathbb{R}^2$ ($r = 0$), since then, we can choose any value of ϕ to describe that point. To define the metric at the origin, one would have to employ another chart. A similar phenomenon arises in the next example.

Example 3.16. Let $M = S^n \subset \mathbb{R}^{n+1}$ be the n-sphere. The first fundamental form on the sphere is called the *standard metric*, g_{std} of S^n. The charts can be described by the stereographic projection from two antipodal points. Let us look at the case $n = 2$, as shown in Figure 3.2. Take these points to be the south pole $SP = (-1,0,0)$ and the north pole $NP = (1,0,0)$. Let $U_1 := S^2 \backslash \{SP\}$, then the chart on U_1 is given by

$$x_1 : U_1 \to \mathbb{R}^2, \quad y = (y^0, y^1, y^2) \mapsto \frac{2}{1+y^0}(y^1, y^2).$$

Let $U_2 := S^2 \backslash \{NP\}$, then the chart on U_2 is given by

$$x_2 : U_1 \to \mathbb{R}^2, \quad y = (y^0, y^1, y^2) \mapsto \frac{2}{1-y^0}(y^1, y^2).$$

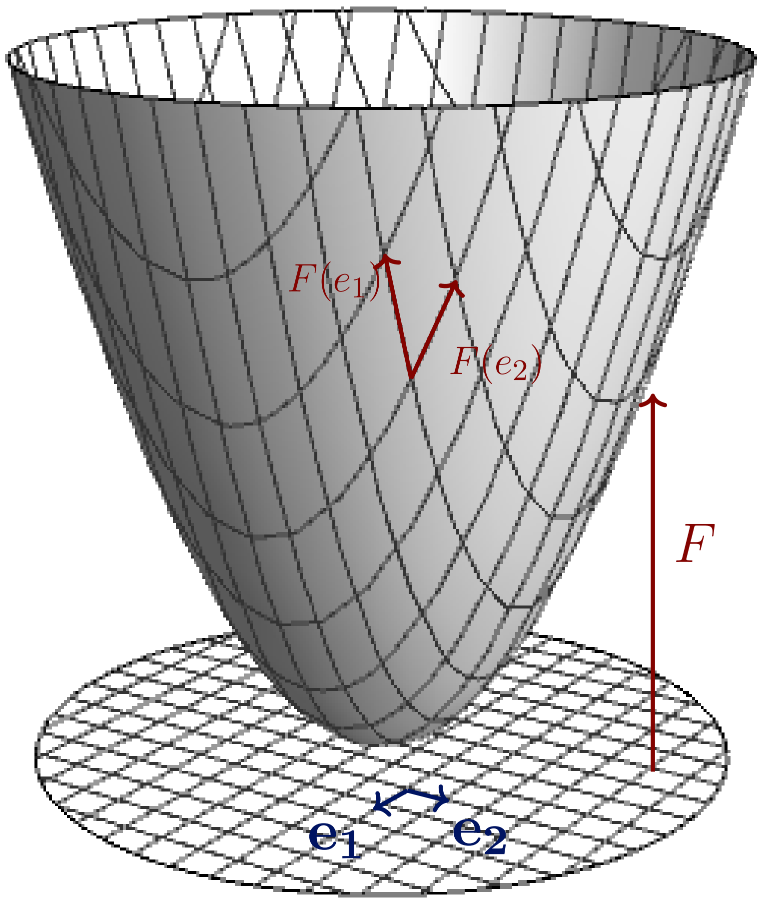

Figure 3.1: The parametrization function F of the paraboloid $P := \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 = 2z\}$ "bending" $\mathbb{R}^2$. As indicated, $\mathbb{R}^2$ is "stretched out" so that the distance between points $F((x_1, y_1))$ and $F((x_2, y_2))$ gets larger if $(x_1, y_1), (x_2, y_2)$ are further away from the origin. In addition, the parametrized basis $\big(F(e_1), F(e_2)\big)$ need not be orthogonal.

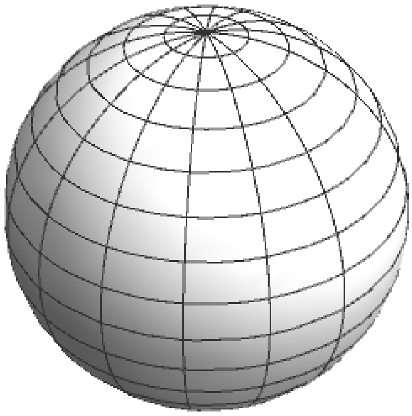

Figure 3.2: Parametrization of the 2-sphere. The standard metric on S^2 is "retracted" near the poles and relatively "stretched out" near the equator.

Essentially, the point on the sphere is projected onto the Euclidean plane tangent on the antipodal point (i.e. NP for x_1 and SP for x_2) along the line through the projection point (SP for x_1, NP for x_2). Both charts are necessary to cover S^2.
Let us express the standard metric in terms of the stereographic projection from the south pole, x_1. Then the inverse is given by

$$F : \mathbb{R}^2 \to S^2 \subset \mathbb{R}^3, \quad x \to \frac{1}{4 + \|x\|^2}\big(4 - \|x\|^2, 4x\big).$$

We compute that

$$\left\langle \frac{\partial F}{\partial x^i}, \frac{\partial F}{\partial x^j} \right\rangle = \frac{16}{(4 + \|x\|^2)^2}\delta_{ij}$$

for all $i, j = 1, 2$. Hence, by Equation (3.5), we obtain

$$(g_{\text{std}})_{ij} = \frac{16}{(4 + \|x\|^2)^2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

As seen in Figure 3.2, the metric on the "equator" of the sphere is relatively stretched out compared to the neighborhood of the poles. The metric confirms this: if $\|x\|$ is large, the point x on the Euclidean plane tangent on the north pole is far away from NP. Hence, the slope of the line through x and SP is smaller and the intersection with S^2 is therefore closer to SP. The expression above tells us that for these points, the metric is "retracted" (i.e. becomes smaller).

The above example using the stereographic projection is useful for illustrating Riemannian metrics, but for S^n it is often convenient to use spherical coordinates. Since the distance from the origin is always 1 for points on the sphere, we can describe points on S^2 using only the polar angle ϕ and azimuthal angle θ. Then, the parametrization function F is given by

$$F(\phi, \theta) = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)$$

so that we can describe the metric $g_{\text{std}}^{\text{sph}}$ with respect to the basis $\big(\frac{\partial}{\partial \phi}, \frac{\partial}{\partial \theta}\big)$. Note that again, we cannot describe SP and NP using this parametrization, since any value of

ϕ can be used to describe these ($\theta = 0$ for SP, $\theta = \pi$ for NP). To be able to describe these points as well, we can repeat this process for a distinct pair of antipodal points to cover S^2.
We compute

$$\left\langle \frac{\partial F}{\partial \phi}, \frac{\partial F}{\partial \phi} \right\rangle = \sin^2(\theta),$$

$$\left\langle \frac{\partial F}{\partial \theta}, \frac{\partial F}{\partial \theta} \right\rangle = \sin^2(\theta) + \cos^2(\theta) = 1,$$

$$\left\langle \frac{\partial F}{\partial \theta}, \frac{\partial F}{\partial \phi} \right\rangle = 0$$

and hence

$$g_{\text{std}}^{\text{sph}}(\phi, \theta) = \begin{pmatrix} \sin^2\theta & 0 \\ 0 & 1 \end{pmatrix}.$$

Noting that $\sin^2\theta \to 0$ as $\theta \to 0$ or $\theta \to \pi$ (i.e. close to the poles) and that $\sin^2\theta = 1$ for $\theta = \pi/2$ (i.e. on the equator), we confirm again that the metric is "retracted" near the poles and relatively "stretched" near the equator.

3.3 Differentiating Vector Fields

Before we talk about the *distance* between two points on a Riemannian manifold, let us define how to differentiate a vector field in a certain direction. This is important for, among other things, defining the divergence of vector fields, which we will need to study the heat equation (see SECTION 4.2). To do this, we require an important type of function called a *Levi-Civita connection*. This will be helpful in defining geodesics hereafter.

We have seen how to differentiate functions on manifolds and we have seen what it means for a vector field to be differentiable (or C^k). But we have not yet seen how to actually differentiate a vector field. What is its differential? Similar to the approach with derivations, we develop an axiomatic formalism to describe differentials. Except, this time, we are differentiating vector fields instead of functions. Let us define the following.

Definition 3.17. Let M be a differentiable manifold and let $k \in \mathbb{N} \cup \{\infty\}$. Let $U \subset M$ open. Then we define the **set of C^k-vector fields defined on** U as $C^k(U, TM)$. Moreover, let $p \in M$. Then, we set

$$\Xi_p := \bigcup_{U \subset M, p \in U} C^\infty(U, TM).$$

Now, the properties of the derivative of a smooth vector field $\eta \in \Xi_p$ along the direction of a tangent vector $\xi \in T_pM$ are incorporated in a kind of function called the Levi-Civita connection, as mentioned before. This function should take η and ξ and give us a derivative in the form of another tangent vector in T_pM, hence it is a function $\Xi_p \times T_pM \to T_pM$.

Definition 3.18. Let (M, g) be a Riemannian manifold. Let $p \in M$. A **Levi-Civita connection** (at p) is a map

$$\nabla : \Xi_p \times T_pM \to T_pM$$

that satisfies the following conditions:

(i) *Locality*: For all $\xi \in T_pM$ and $\eta \in C^\infty(U, TM)$ and for all $\tilde{U} \subset U$ with $p \in \tilde{U}$ we must have
$$\nabla_\xi \eta = \nabla_\xi(\eta|_{\tilde{U}}).$$

(ii) *Linearity in first argument*: For all $\xi_1, \xi_2 \in T_pM$, for all $a, b \in \mathbb{R}$ and for all $\eta \in \Xi_p$ we have
$$\nabla_{a\xi_1 + b\xi_2}\eta = a\nabla_{\xi_1}\eta + b\nabla_{\xi_2}\eta.$$

(iii) *Additivity in the second argument*: For all $\xi \in T_pM$ and for all $\eta_1, \eta_2 \in \Xi_p$ we must have
$$\nabla_\xi(\eta_1 + \eta_2) = \nabla_\xi\eta_1 + \nabla_\xi\eta_2.$$

(iv) *Product Rule I*: For all $f \in C_p^\infty$, for all $\eta \in \Xi_p$ and for all $\xi \in T_pM$ we demand that
$$\nabla_\xi(f \cdot \eta) = \partial.\eta|_p + f(p) \cdot \nabla_\xi\eta.$$

(v) *Product Rule II*: For all $\xi \in T_pM$ and for all $\eta_1, \eta_2 \in \Xi_p$:
$$\partial_\xi g(\eta_1, \eta_2) = g|_p(\nabla_\xi\eta_1, \eta_2|_p) + g|_p(\eta_1|_p, \nabla_\xi\eta_2).$$

(vi) *Torsion-freeness*: For all charts $x : U \to V$ of M with $p \in U$ we have
$$\nabla_{\frac{\partial}{\partial x^i}|_p}\frac{\partial}{\partial x^j} = \nabla_{\frac{\partial}{\partial x^j}|_p}\frac{\partial}{\partial x^i}$$
for all i, j, which is similar to how the curl of a gradient of a usual vector field always vanishes.

We will not show this here, but (iii) and (iv) imply $\mathbb{R}$-linearity in the second argument. Moreover, (vi) holds for any chart containing p (if it holds for one chart).
Instead of starting from such a function, sometimes it is necessary to construct one instead.

Definition 3.19. Let $x : U \to V$ be a chart. We write $\nabla_{\frac{\partial}{\partial x^i}|_p}\frac{\partial}{\partial x^j} \in T_pM$ as a tangent vector as follows:
$$\nabla_{\frac{\partial}{\partial x^i}|_p}\frac{\partial}{\partial x^j} = \Gamma_{ij}^k\big(x(p)\big) \cdot \frac{\partial}{\partial x^k}\bigg|_p, \tag{3.6}$$
where the Γ_{ij}^k are named **Christoffel Symbols**.

Lemma 3.20. The Christoffel symbols determine ∇.

Proof. Let $\xi = \xi^i \frac{\partial}{\partial x^i}\big|_p \in T_pM$ and let $\eta = \eta^i \frac{\partial}{\partial x^i}\big|_p \in \Xi_p$. We denote the properties of ∇ similar to Definition 3.18. We have
$$\begin{aligned}\nabla_\xi(\eta) &\overset{\text{(ii), (iii)}}{=} \xi^i \nabla_{\frac{\partial}{\partial x^i}|_p}\left(\eta^j \frac{\partial}{\partial x^j}\right) \\ &\overset{(iv)}{=} \xi^i \left(\frac{\partial \eta^j}{\partial x^i}\bigg|_{x(p)} \cdot \frac{\partial}{\partial x^j}\bigg|_p + \eta^j|_{x(p)} \cdot \nabla_{\frac{\partial}{\partial x^i}|_p}\frac{\partial}{\partial x^j}\right) \\ &\overset{(3.6)}{=} \xi^i \left(\frac{\partial \eta^j}{\partial x^i}\bigg|_{x(p)} + \big(\eta^j|_{x(p)} \cdot \Gamma_{ij}^k(x(p))\big)\right)\frac{\partial}{\partial x^k}\bigg|_p.\end{aligned} \tag{3.7}$$

□

Remark. Note that (vi) is equivalent to symmetry of the Christoffel symbols in the lower indices, i.e. $\Gamma^k_{ij} = \Gamma^k_{ji}$ for all i, j, k.

We are almost ready to determine the definition of a differential of a vector field. However, we have to ask ourselfs: how can we be sure that a Levi-Civita connection even exists at all points p? And is it unique, or are there multiple? As we will see shortly, there must exist exactly one Levi-Civita connection at all points of a Riemannian Manifold. The proof of this is rather involved, but is worth going through, since we will find an expression for the Christoffel symbols as a function of the Riemannian metric matrix elements (g_{ij}) along the way.

Proposition 3.21. (The Fundamental Theorem of Riemannian Geometry). Let (M, g) be a Riemannian manifold and let $p \in M$. Then there exists a Levi-Civita connection at p and it is unique.

Proof. First, we show uniqueness. Let $x : U \to V$ be a chart of M with $p \in U$. We are going to differentiate the metric function with respect to components of x. We find (where we implicitly evaluate in p):

$$\begin{aligned}
\frac{\partial g_{ij}}{\partial x^k} &= \frac{\partial}{\partial x^k} g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) \\
&\overset{\text{(v)}}{=} g\left(\nabla_{\frac{\partial}{\partial x^k}} \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) + g\left(\frac{\partial}{\partial x^i}, \nabla_{\frac{\partial}{\partial x^k}} \frac{\partial}{\partial x^j}\right) \\
&= g\left(\Gamma^l_{ki} \frac{\partial}{\partial x^l}, \frac{\partial}{\partial x^j}\right) + g\left(\frac{\partial}{\partial x^i}, \Gamma^l_{kj} \frac{\partial}{\partial x^j}\right) \\
&\overset{*}{=} \Gamma^l_{ki} g\left(\frac{\partial}{\partial x^l}, \frac{\partial}{\partial x^i} + \Gamma^l_{kj} g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^l}\right)\right. \\
&= \Gamma^l_{ki} g_{lj} + \Gamma^l_{kj} g_{il}
\end{aligned}$$

where we used the linearity of g at $*$. Summarizing this result and rearranging the indices, we obtain the following set of equations:

$$\frac{\partial g_{ij}}{\partial x^k} = \Gamma^l_{ki} g_{lj} + \Gamma^l_{kj} g_{il} \tag{3.8}$$

$$\frac{\partial g_{ik}}{\partial x^j} = \Gamma^l_{ji} g_{lk} + \Gamma^l_{jk} g_{il} \tag{3.9}$$

$$\frac{\partial g_{kj}}{\partial x^i} = \Gamma^l_{ik} g_{lj} + \Gamma^l_{ij} g_{kl}. \tag{3.10}$$

Now, since we are discussing a Levi-Civita connection which, in particular, satisfies the torsion-freeness, we may assume symmetry of the Christoffel symbols in the lower indices. Using this while writing out (3.8)−(3.9)+(3.10) we find

$$\frac{\partial g_{ij}}{\partial x^k} - \frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{kj}}{\partial x^i} = 2\Gamma^l_{ki} g_{lj}.$$

Denote $(g^{ij})_{ij}$ be the inverse of $(g_{ij})_{ij}$ so that $g^{ij} g_{jk} = \delta^i_k$ (note that this inverse exist as g is non-degenerate, meaning all $g|_p$ are). We obtain now

$$\left(\frac{\partial g_{ij}}{\partial x^k} - \frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{kj}}{\partial x^i}\right) g^{jm} = 2\Gamma^l_{ki} g_{lj} g^{jm} = 2\Gamma^l_{ki} \delta^m_l = 2\Gamma^m_{ki}.$$

Renaming the indices as $(k \to j, m \to k, j \to m)$ we find the following important relation:

$$\Gamma^k_{ij} = \frac{1}{2} g^{mk} \left(\frac{\partial g_{im}}{\partial x^j} + \frac{\partial g_{jm}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^m} \right). \tag{3.11}$$

As a result, the Christoffel symbols are uniquely determined, smooth functions $V \to \mathbb{R}$ for any chart $x : U \to V$ on (M, g). Hence, ∇ is uniquely determined by the components of g and their derivatives with respect to the components of the chart x.

We are yet to show the existence of the Levi-Civita connection at p. Define Γ^k_{ij} as in Equation (3.11) and define ∇ as in Equation (3.7). Then (i) follows from the locality of derivations $\frac{\partial}{\partial x^i}$, (ii) follows from the linearity of g, (iii) follows immediately from the linearity in η^j in the expression in (3.7) and (vi) follows from the fact that we can swap indices i and j in the expression between the brackets in (3.11) while only changing the order of addition (that is, doing nothing). Let us show that ∇ obeys the first product rule (iv):

$$\begin{aligned}
\nabla_\xi(f\eta) &= \xi^i \left(\frac{\partial (f \cdot \eta^k)}{\partial x^i} + f \eta^j \Gamma^k_{ij} \right) \frac{\partial}{\partial x^k} \\
&\overset{*}{=} f \cdot \xi^i \left(\frac{\partial \eta^k}{\partial x^i} + \eta^j \Gamma^k_{ij} \right) \frac{\partial}{\partial x^k} + \xi^i \frac{\partial f}{\partial x^i} \eta^k \frac{\partial}{\partial x^k} \\
&= f \cdot \nabla_\xi \eta + \partial_\xi f \cdot \eta.
\end{aligned}$$

where we used the usual product rule in $*$. Finally, we check the second product rule (v). In the following, we rename indices and use the fact that $g_{ij} g^{ki} = \delta^k_j$ occasionally. We write out

$$\begin{aligned}
&\partial_\xi g(\xi, \eta) - g(\nabla_\zeta \xi, \eta) - g(\xi, \nabla_\zeta \eta) \\
&\quad = \zeta^k \frac{\partial}{\partial x^k} (g_{ij} \xi^i \eta^j) - g_{ij} \zeta^k \left(\frac{\partial \xi^i}{\partial x^k} + \xi^l \Gamma^i_{lk} \right) \eta^j - g_{ij} \xi^i \zeta^k \left(\frac{\partial \eta^j}{\partial x^k} + \eta^l \Gamma^j_{lk} \right) \\
&\quad = \zeta^k \frac{\partial g_{ij}}{\partial x^k} \xi^i \eta^j - g_{ij} \zeta^k \xi^l \Gamma^i_{lk} \eta^j - g_{ij} \xi^{ik} \eta^l \Gamma^j_{lk} \\
&\quad = \xi^i \eta^j \zeta^k \big(\frac{\partial g_{ij}}{\partial x^k} - g_{lj} \Gamma^l_{ik} - g_{il} \Gamma^l_{jk} \big) \\
&\quad \overset{(3.11)}{=} \xi^i \eta^j \zeta^k \Bigg(\frac{\partial g_{ij}}{\partial x^k} - \frac{1}{2} g_{lj} g^{ml} \left(\frac{\partial g_{im}}{\partial x^k} + \frac{\partial g_{km}}{\partial x^i} - \frac{\partial g_{ik}}{\partial x^m} \right) \\
&\qquad - \frac{1}{2} g_{il} g^{ml} \left(\frac{\partial g_{jm}}{\partial x^k} + \frac{\partial g_{km}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^m} \right) \Bigg) \\
&\quad = \xi^i \eta^j \zeta^k \left(\frac{\partial g_{ij}}{\partial x^k} - \frac{1}{2} \left(\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ik}}{\partial x^j} \right) - \frac{1}{2} \left(\frac{\partial g_{ji}}{\partial x^k} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^i} \right) \right) \\
&\quad = 0.
\end{aligned}$$

This concludes the proof. □

Note that the differential of a vector field (or actually, the Levi-Civita connection ∇) depends on the metric g (specifically the second product rule (v), which is therefore sometimes referred to as the *compatibility with the metric*). It is indeed quite curious that both differentiability and the derivative of a function on a differentiable manifold are defined, as well as the differentiability of vector fields, while on the other hand,

we need a Riemannian metric to define the derivative of a vector field.
Let us now formally define the derivative of a vector field.

Definition 3.22. Let (M, g) be a Riemannian manifold and let ∇ be its Levi-Civita connection. Moreover, let $p \in M$, $\xi \in T_pM$ and $\eta \in \Xi_p$. We call

$$\nabla_\xi \eta \in T_pM$$

the **covariant derivative** of η in direction ξ.

Let us look at an example of the covariant derivative and Christoffel symbols for the Euclidean plane similar to Examples 2.28 and 3.11.

Example 3.23. Let $(M, g) = (\mathbb{R}^2, g_{\text{Eucl}})$. In Cartesian coordinates (i.e. components of chart $x = \text{id}$) x^1 and x^2, the $g_{ij} = \delta_{ij}$ are constant, hence its derivatives are zero and we find $\Gamma^k_{ij} = 0$ for all $i, j, k = 1, 2$. Let us look at the derivative of the vector field $\frac{\partial}{\partial\phi}$ along direction $\frac{\partial}{\partial\phi}$ in polar coordinates r, ϕ.

$$\begin{aligned}
\nabla_{\frac{\partial}{\partial\phi}} \frac{\partial}{\partial\phi} &= \nabla_{-x^2\frac{\partial}{\partial x^1}+x^1\frac{\partial}{\partial x^2}} \left(-x^2\frac{\partial}{\partial x^1} + x^1\frac{\partial}{\partial x^2} \right) \\
&= -x^2\frac{\partial}{\partial x^1} + x^1\frac{\partial}{\partial x^2}(-x^2)\frac{\partial}{\partial x^1} + -x^2\frac{\partial}{\partial x^1} + x^1\frac{\partial}{\partial x^2}(x^1)\frac{\partial}{\partial x^2} \\
&= -x^1\frac{\partial}{\partial x^1} - x^2\frac{\partial}{\partial x^2} = -r\frac{\partial}{\partial r}.
\end{aligned}$$

That is, the derivative of the vector field depicted in Figure 2.1(b) has magnitude r and points radially inward (this can be compared to centripetal force/acceleration in physics where the derivative w.r.t. angle ϕ corresponds to angular momentum/velocity).

Remember that, in polar coordinates, the metric is given by

$$(g^{\text{polar}}_{\text{Eucl}})_{ij}(r, \phi) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}$$

and similarly

$$(g^{\text{polar}}_{\text{Eucl}})^{ij}(r, \phi) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{r^2} \end{pmatrix}.$$

Thus, the Christoffel symbols with respect to polar coordinates, computed with Equation (3.11), are given by

$$\Gamma^1_{11} = \frac{1}{2}\big(1 \cdot (0 + 0 - 0) + 0 \cdot \ldots\big) = 0$$

and similarly

$$\Gamma^2_{11} = \Gamma^1_{12} = \Gamma^1_{21} = \Gamma^2_{22} = \Gamma^1_{11} = 0.$$

Finally,

$$\Gamma^2_{12} = \Gamma^2_{21} = \frac{1}{2}\left(\frac{1}{r^2}\left(\frac{\partial g_{12}}{\partial\phi} + \frac{\partial g_{22}}{\partial r} - \frac{\partial g_{21}}{\partial\phi}\right) + 0 \cdot \ldots\right) = \frac{1}{r}$$

and similarly, $\Gamma^1_{22} = -r$. The Christoffel symbols allow us to directly compute through Equation (3.7)

$$\nabla_{\frac{\partial}{\partial\phi}} \frac{\partial}{\partial\phi} = \Gamma^1_{22}\frac{\partial}{\partial r} + \Gamma^2_{22} = -r\frac{\partial}{\partial r}$$

which is indeed the same result as in Cartesian coordinates treated above.

Other than defining a derivative of a vector field in a certain direction $\xi \in T_pM$, we can also define the derivative of a vector field *along a map*.

Definition 3.24. Let M and N be differentiable manifolds and let $f : N \to M$ be a map. Then a map $\zeta : N \to TM$ is a **vector field along** f if

$$\pi_M \circ \zeta = f$$

where $\pi_M : TM \to M$ is the "projection map" defined for vector fields in differential geometry.

Example 3.25. (Vector fields along curves). Let $N = I \subset \mathbb{R}$ be an open interval and $f = c : I \to M$ be a curve. Then ζ as in Definition 3.24 is a vector field along the curve c. A special case which is important for the notion of geodesics is $\zeta(t) = \dot{c}(t) := c_t^0$ where $c_t(s) := c(t+s)$. This is called the **velocity field** of c.

We can also define a covariant derivative along a map.

Definition 3.26. Let N be a differentiable manifold. Let (M, g) be a Riemannian manifold. Furthermore, let $f : N \to M$ be a differentiable map and let $\eta : N \to TM$ be a differentiable vector field along f. Let $p \in N$ and $\xi \in T_pN$. Then, we define the **covariant derivative** $\nabla_\xi \eta \in T_{f(p)}M$ as follows:
Let $x : U \to V$ be any chart of (M, g) with $f(p) \in U$, and write for points $q \in N$, as usual,

$$\eta(q) = \eta^j(q) \frac{\partial}{\partial x^j}\bigg|_{f(p)}$$

with differentiable coefficient functions $\eta^1, \ldots, \eta^n$ that are well-defined on $f^{-1}(U)$. Now, choose a curve $c : (-\epsilon, \epsilon) \to N$ with associated equivalence class $c^0 = \xi$ and define $\nabla_\xi \eta$ as in Equation (3.7), while replacing $p \to f(p)$ and differentiating along the curve rather than w.r.t. the components of the chart x. In other words,

$$\nabla_\xi \eta = \left(\partial_\xi \eta^k + \eta^j(p) df^i(\xi) \Gamma^k_{ij}(x(f(p))) \right) \frac{\partial}{\partial x^k}\bigg|_{f(p)}.$$

Proposition 3.27. Let the setting be the same as Definition 3.26. Then the covariant derivative $\nabla_\xi \eta$ is defined independently of charts x and the choice of curve c to represent $c^0 = \xi$. Moreover, if we have that whenever η is of the form $\zeta \circ f$ for some differentiable vector field on M ζ, then it implies that

$$\nabla_\xi \eta = \nabla_{df|_p(\xi)} \zeta$$

then the operator ∇ is the Levi-Civita connection.

Proof. This follows directly from the fact that the changes made to Equation (3.7) in Definition **??** coincide with the fact that $\nabla_\xi \eta = \nabla_{df|_p(\xi)} \zeta$. □

Notation 3.28. Let the setting be the same as the Definition above. Let y be a chart on N containing p (we call this *local coordinates* y). Then, we write

$$\frac{\nabla \eta}{\partial y^i}(p) := \nabla_{\frac{\partial}{\partial y^i}} \eta = \left(\frac{\partial \eta^k}{\partial y^l}\bigg|_{y(p)} + \frac{\partial f^i}{\partial y^l}(p) \cdot \eta^j(y(p)) \cdot \Gamma^k_{ij}(x(f(p))) \right) \frac{\partial}{\partial x^k}(p). \quad (3.12)$$

If dim $N = 1$, we write

$$\frac{\nabla \eta}{\partial t} =: \frac{\nabla \eta}{dt}.$$

Example 3.29. Let $c : I \to M$ be a continuous piecewise C^2-curve and let c_t^0 be the velocity field. We have

$$\frac{\nabla \dot{c}}{dt}(t) = \left((\ddot{c})^k(t) + \dot{c}^i(t) \cdot \dot{c}^j(t) \cdot \Gamma_{ij}^k\big(x(c(t))\big) \right) \frac{\partial}{\partial x^k}\bigg|_{c(t)} \tag{3.13}$$

where $c^k := x^k \circ c$. This can be interpreted as a kind of vector field that assigns an "acceleration" (i.e. a change in velocity) to each point in the image of c. This notion will be important for defining geodesics.

3.4 Geodesics

The last phenomenon we need to explore before being able to complete a formalism of Riemannian geometry using a distance function is the notion of the paths of least energy between two points on a Riemannian Manifold, *geodesics*. Let us define the energy of a curve.

Definition 3.30. Let (M, g) be a Riemannian manifold. Let $c : [a, b] \to M$ be a continuous C^1-curve. Then the **energy** of c is given by

$$E[c] := \frac{1}{2} \int_a^b g\big(\dot{c}(t), \dot{c}(t)\big) dt.$$

Note that since g is positive definite we have $E[c] \geq 0$ where the equality holds only if c is a constant path, i.e. $\dot{c} = 0$. In light of the hints towards the definition of geodesics throughout this thesis, we can ask the question: are there curves with minimal energy between two given points?
To answer this question, we define first the notion of a *variation* of a curve c, which is essentially a "continuous film" of different (but similar) curves around c in the same sense that a homotopy in multivariable analysis or topology is a "film" of paths. One could also interpret a variation as a "curve of curves", i.e. a curve where the points on the curve represent a unique curve.

Definition 3.31. Let M be a differentiable manifold and let $c : [a, b] \to M$ be a smooth curve. Then the smooth map

$$C : (-\epsilon, \epsilon) \times [a, b] \to M$$

with $C(0, t) = c(t)$ for all $t \in [a, b]$ is called a **variation** of c. If, in addition, $C(s, a) = c(a)$ and $C(s, b) = c(b)$ for all $s \in (-\epsilon, \epsilon)$, we call $C(s, t)$ a **variation with fixed endpoints**.
The vector field $\eta(t) := \frac{\partial C}{\partial s}(0, t)$ is called the **variational vector field**.

For a variation with fixed endpoints, $C(s, a)$ and $C(s, b)$ are constant functions (with respect to s), hence the variational vector field vanishes at a and b.

Notation 3.32. We often write $C(s, t) =: C_s(t)$.

If we have such a variation, we can ask ourselves how the energy of all the curves in the variation "varies" as a function of s, or particularly, what the "slope" (i.e. the change rate of the energy with respect to s) of this energy is on the curve c. We call this "slope" the first variation of the energy.

Theorem 3.33. *Let (M, g) be a Riemannian manifold. Let $c : [a, b] \to M$ be a smooth curve and let $C : (-\epsilon, \epsilon) \times [a, b] \to M$ be a variation of the curve. Denote the variational vector field by η. Then,*

$$\frac{d}{ds}E[C_s]\big|_{s=0} = -\int_a^b g\Big(\eta(t), \frac{\nabla}{dt}\dot{c}(t)\Big)dt + g\big(\eta(b), \dot{c}(b)\big) - g\big(\eta(a), \dot{c}(a)\big).$$

Proof. First, notice that the "velocity field" of the variation C is the field $\dot{C}_s(t) = \frac{\partial C}{\partial t}(s, t)$. Let us prove the theorem with a direct computation, where the roman numbers denote the properties of the Levi-Civita connection from Definition 3.18.

$$\begin{aligned}
\frac{d}{ds}E[C_s]\big|_{s=0} &= \frac{1}{2}\frac{d}{ds}\Big|_{s=0}\int_a^b g\big(\dot{C}_s(t), \dot{C}_s(t)\big)dt \\
&\overset{*}{=} \frac{1}{2}\int_a^b \frac{\partial}{\partial s}\Big|_{s=0} g\Big(\frac{\partial C}{\partial t}(s,t), \frac{\partial C}{\partial t}(s,t)\Big)dt \\
&\overset{\text{(v)}}{=} \Big[g\Big(\frac{\nabla}{\partial s}\frac{\partial C}{\partial t}(0,t), \frac{\partial C}{\partial t}(0,t)\Big) + g\Big(\frac{\partial C}{\partial t}(0,t), \frac{\nabla}{\partial s}\frac{\partial C}{\partial t}(0,t)\Big)\Big]dt \\
&= \int_a^b g\Big(\frac{\nabla}{\partial s}\frac{\partial C}{\partial t}(0,t), \frac{\partial C}{\partial t}(0,t)\Big)dt \\
&\overset{\text{(vi)}}{=} \int_a^b g\Big(\frac{\nabla}{\partial t}\frac{\partial C}{\partial s}(0,t), \frac{\partial C}{\partial t}(0,t)\Big)dt \\
&= \int_a^b g\Big(\frac{\nabla}{dt}\eta(t), \dot{c}(t)\Big)dt \\
&\overset{\text{(v)}}{=} \int_a^b \Big[\frac{d}{dt}g\big(\eta(t), \dot{c}(t)\big) - g\Big(\eta(t), \frac{\nabla}{dt}\dot{c}(t)\Big)\Big]dt
\end{aligned}$$

and the result follows from integrating the first term using the fundamental theorem of calculus. In the above, at $*$ we used the well-known result in multivariable calculus that limits (and thus derivatives) of integrals are equal to the integral of the limit of the integrand as long as the integration- and limit parameters are independent. Moreover, the torsion-freeness at (vi) allows us to switch parameters s and t in the expression $\frac{\nabla}{\partial s}\frac{\partial C}{\partial t}$. □

Collorary 3.34. If C is a variation with fixed endpoints, we have that the boundary terms in the variation of the energy vanish (as $\eta(b) = \eta(a) = 0$) and we have

$$\frac{d}{ds}E[C_s]\big|_{s=0} = -\int_a^b g\Big(\eta(t), \frac{\nabla}{dt}\dot{c}(t)\Big)dt.$$

Let us show that given a smooth curve c and a vector field η along it, there always exists a variation C so that its variational vector field is η.

Lemma 3.35. Let $c : [a, b] \to M$ be a smooth curve and let η be a vector field along c. There exists a variation C of c with variational vector field η. Moreover, if $\eta(a) = \eta(b) = 0$, we can choose the variation to have fixed endpoints.

Proof. Let $x : U \to V$ be a chart of M. Assume that $c(t) \in U$ if $\eta(t) \neq 0$. Writing $\eta = \eta^j \frac{\partial}{\partial x^j}$ we set the variation

$$C(s,t) := \begin{cases} x^{-1}\Big(\big(c^1(t), \ldots, c^n(t)\big) + s\big(\eta^1(t), \ldots, \eta^n(t)\big)\Big), & c(t) \in U \\ c(t), & c(t) \notin U. \end{cases}$$

Then, the associated vector field satisfies

$$\begin{aligned} \left(\frac{\partial C}{\partial s}(0,t)\right)^j &= dx^j\left(\frac{\partial C}{\partial s}(0,t)\right) \\ &= \frac{\partial (x^j \circ C)}{\partial s}(0,t) \\ &= \frac{\partial (c^j(t) + s\eta^j(t))}{\partial s}\bigg|_{s=0} \end{aligned}$$

which is precisely $\eta^j(t)$ (by simply differentiating). Thus, C has variational vector field η and if $\eta(a) = \eta(b) = 0$ we have that C has fixed endpoints.
To extend this argument to the general case of all M instead of an open subset U, we use the compactness of the interval $[a,b]$ and the continuity of c to find that the set $c([a,b])$ is compact. Hence, we can find a finite subcover of opens that are domains of charts. This way, the variation can be constructed piecewise. □

Now we are ready to look at geodesics. In fact, it follows from Theorem 3.33 and Lemma 3.35 that an energy-minimizing curve is precisely a curve that "feels no acceleration", i.e. the "acceleration field" defined in Example 3.29 vanishes on the image of the curve. The latter is indeed the definition of a geodesic. First, let us introduce some useful notation.

Notation 3.36. Let M be a differentiable manifold. Let $p, q \in M$. We define the **set of smooth curves from p to q** by

$$O_{p,q}(M) := \big\{\text{smooth curves } c : [a,b] \to M \mid c(a) = p,\ c(b) = q\big\}.$$

Collorary 3.37. For a Riemannian manifold (M, g) and curve $c \in O_{p,q}(M)$ we have that the following statements are equivalent:

(i) The curve is an "extreme" of the variational energy. This means

$$\frac{d}{ds}E[C_s]\big|_{s=0} = 0$$

for all variations C_s of c with fixed endpoints.

(ii) For all t the "acceleration field" vanishes, i.e.

$$\frac{\nabla}{dt}\dot{c}(t) = 0.$$

Proof. The implication "(ii) $\Longrightarrow$ (i) is obvious, just fill in $\frac{\nabla}{dt}\dot{c}(t) = 0$ in the expression in Corollary 3.34. For the converse, let c be defined on $[a,b]$. Assume (ii) is not true, i.e. there is a $\tau \in (a,b)$ so that $\frac{\nabla}{dt}\dot{c}(\tau) \neq 0$. We show that (i) must be false too. Under the assumption, there must exist a tangent vector $\xi \in T_{c(\tau)}M$ with

$$g\left(\xi, \frac{\nabla}{dt}\dot{c}(\tau)\right) > 0$$

(by non-degenerateness of g). Let η' be the vector field parallel to c so that $\eta'(\tau) = \xi$. By the definition of continuity, there must be a neighborhood of τ so that the metric does not vanish, i.e. there exists an $\epsilon > 0$ so that $I_\epsilon := (\tau - \epsilon, \tau + \epsilon) \subset [a, b]$ and

$$g\left(\eta'(t), \frac{\nabla}{dt}\dot{c}(\tau)\right) > 0$$

for all $t \in I_\epsilon$. Next, choose a smooth function $f : [a, b] \to \mathbb{R}$ with $f(t) > 0$ for all $t \in I_\epsilon$ and $f(t) = 0$ for all other t (compare this to a partition of unity in topology). Now set $\eta(t) = f(t) \cdot \eta'(t)$. Then we have

$$g\left(\eta(t), \frac{\nabla}{dt}\dot{c}(\tau)\right) = f(t) \cdot g\left(\eta'(t), \frac{\nabla}{dt}\dot{c}(\tau)\right)$$

which is positive on I_ϵ and zero on $[a, b] \backslash I_\epsilon$. By Lemma 3.35 we can chose a variation with fixed endpoints C of c so that the variational vector field is η. Then, the variation of the energy for this variation is non-zero, as

$$\frac{d}{ds} E[C_s]\big|_{s=0} = -\int_a^b g\left(\eta(t), \frac{\nabla}{dt}\dot{c}(\tau)\right) dt < 0$$

and we may conclude that $\frac{\nabla}{dt}\dot{c} = 0$ on (a, b). Since this is a continuous map, it must also be 0 on boundary points a and b. □

Definition 3.38. A **geodesic** is a smooth curve c such that $\frac{\nabla}{dt}\dot{c} = 0$.

From a physics point of view, this can be interpreted quite intuitively. A geodesic between two points is a curve that feels no exterior force (thus no acceleration), or "friction". Since, in physics, force is considered to be the spatial derivative of energy (in fact, the negative gradient of the energy, where exterior force is considered negative). Since the exterior force is zero, the energy must be constant along the geodesic. If there had been any force, the energy would be higher, from which we can conclude that the geodesic minimizes the energy of curves between two points.

Example 3.39. Let $(M, g) = (\mathbb{R}^n, g_{\text{Eucl}})$. In Cartesian coordinates, we see that $\frac{\nabla}{dt}\dot{c} = 0$ is equivalent with the fact that $\ddot{c} = 0$, since the Christoffel symbols vanish (apply Equation (3.13)). Solving this "differential equation", we get that $c(t) = p + tv$ so that geodesics are straight lines (parametrized with constant speed v) from p to $q = p + bv$.

Lemma 3.40. For any geodesic c, the norm induced by the Riemannian metric of the tangent vectors $\dot{c}$ is constant.

Proof. By the second product rule of the Levi-Civita connection (v) in Definition 3.18) we find

$$\frac{d}{dt} g(\dot{c}, \dot{c}) = 2g\left(\frac{\nabla}{dt}\dot{c}, \dot{c}\right) = 0$$

as $\frac{\nabla}{dt}\dot{c} = 0$ for geodesics and g is non-degenerate. □

A logical step to take is to show that geodesics always exist and are unique.

Proposition 3.41. Let (M, g) be a Riemannian manifold and let $p \in M$ and $\xi \in T_pM$. Then, there exists an open interval $I \in \mathbb{R}$ with $0 \in I$ and a geodesic $c : I \to M$ with $c(0) = p$ and $\dot{c}(0) = c^0 = \xi$.
Moreover, if $c : I \to M$ and $c' : I' \to M$ are two such geodesics with $c(0) = c'(0)$ and $\dot{c}(0) = \dot{c}'(0)$, they must agree on their common domain $I \cap I'$.

Proof. Let $x : U \to V$ be a chart of M so that $p \in U$. By Equation (3.13) we see that c is a geodesic (i.e. $\frac{\nabla}{dt}\dot{c} = 0$) if and only if, for all $k = 1, 2, \dots, n$, we have

$$(\ddot{c})^k + \Gamma^k_{ij}\big(c^1, \dots, c^n\big) \cdot \dot{c}^i \cdot \dot{c}^j = 0$$

with $c^k = x^k \circ c$. This is a system of ordinary differential equations (of the second order). Thus, by the existence and uniqueness theorem of differential equations (also known as the Theorem of Picard-Lindelöf), we see that c exists and is unique up to the assertions above. □

Let us look at an example: the geodesics on (S^n, g_{std}).

Example 3.42. We can predict that the geodesic between two points on the sphere lies on the great circle through them. If c is a geodesic, then $g_{\text{std}}(\dot{c}, \dot{c}) = a > 0$ (since it is constant and ≥ 0, while equality holds only if c is the constant curve). Thus, we seek a parametrization of this great circle as

$$c(t) := p \cdot \cos(at) + \frac{\Phi_p(\xi)}{\|\Phi_p(\xi)\|} \sin(at)$$

so that it satisfies the initial conditions

$$c(0) = p,$$

$$\dot{c}(0) = \xi, \quad \iff \quad \frac{dc}{dt}(0) = \Phi_p(\xi).$$

In our ansatz we find

$$\frac{dc}{dt}(0) = \frac{\Phi_p(\xi)}{\|\Phi_p(\xi)\|} \cdot a$$

thus we must have that $a = \|\Phi_p(\xi)\| = \|\xi\|$. Therefore, we conclude that the unique geodesic on the sphere at p along ξ is

$$c(t) = p \cdot \cos\big(\|\xi\| t\big) + \frac{\Phi_p(\xi)}{\|\xi\|} \cdot \sin\big(\|\xi\| t\big). \tag{3.14}$$

We continue the discussion on geodesics by introducing the Riemannian exponential function. A useful insight, before introduction of this function, is that if c_ξ is a geodesic with $c(0) = p \in M$ and $\dot{c}(0) = \xi \in T_pM$, then $c'(t) := c_\xi(at)$ is also a geodesic for all $a \in \mathbb{R}$ (given that c is defined on all of $\mathbb{R}$). Indeed, by the usual product rule,

$$\frac{\nabla}{dt}\dot{c}'(t) = \frac{\nabla}{dt}\big(a \cdot \dot{c}_\xi(at)\big) = a^2\left(\frac{\nabla}{dt}\dot{c}_\xi\right)(at) = 0.$$

Since we have that $c'(0) = c_\xi(0) = p$ and $\dot{c}'(0) = a \cdot \dot{c}_\xi(0) = a\xi$, we conclude that $c' = c_{a\xi}$. Notice especially that $c_\xi(a) = c_{a\xi}(1)$.

Definition 3.43. Let (M, g) be a Riemannian manifold. Let $p \in M$ and $\xi \in T_pM$. Let

$$D_p := \big\{\xi \in T_pM \mid \text{the maximal domain of } c_\xi \text{ contains } 1\big\}.$$

Then the **Riemannian exponential map** $\exp_p : D_p \to M$ is defined by

$$\exp_p(\xi) := c_\xi(1).$$

Note that, for any $p \in M$, we have that $\exp_p(0) = p$ since c_0 is the constant curve at p ($c_0(t) := p$ for all t).

Lemma 3.44. The curve $c(t) := \exp_p(t\xi)$ is the unique geodesic with initial values $p \in M$ and $\xi \in T_pM$.

Proof. By the discussion before the definition of the Riemannian exponential map, we have $\exp_p(t\xi) = c_{t\xi}(1) = c_\xi(t)$. □

Example 3.45. (1) Let $(M, g) = (\mathbb{R}^n, g_{\text{Eucl}})$. Then $D_p = T_p\mathbb{R}^n$ since all curves are defined at 1. We have

$$\exp_p(\xi) = p + 1 \cdot \Phi_p(\xi) = p + \Phi(\xi)$$

for canonical isomorphism Φ_p.

(2) Let $(M, g) = (S^n, g_{\text{std}})$. Similar to (1), here we have $D_p = T_pM$ and by Example 3.42 we have

$$\exp_p(\xi) = p \cdot \cos\left(\|\xi\|\right) + \frac{\Phi_p(\xi)}{\|\xi\|} \cdot \sin\left(\|\xi\|\right).$$

Lemma 3.46. The differential of the Riemannian exponential map at 0 is given by the canonical isomorphism

$$d\exp_p\big|_0 = \Phi_0 : T_0D_p = T_0(T_pM) \to T_pM.$$

Proof. Let $\xi \in T_pM$. We compute

$$d\exp_p\big|_0(\Phi_0^{-1}(\xi)) = d\exp_p\left(\frac{d}{dt}(t\xi)\big|_{t=0}\right) = \frac{d}{dt}\exp_p(t\xi)\big|_{t=0} = \xi.$$

Hence $d\exp_p\big|_0 \circ \Phi_0^{-1} = \text{id}$. □

In particular, we have that $d\exp_p\big|_0$ is invertible.

Collorary 3.47. Let $p \in M$. There exists an open neighborhood $W_p \subset D_p \subset T_pM$ of 0 such that

$$\exp_p\big|_{W_p} : W_p \to \exp_p(W_p) =: U_p$$

is a diffeomorphism.

Proof. Since $d\exp_p\big|_0$ is invertible, the assertion follows from the inverse function theorem 2.13. □

Collorary 3.48. Let $p \in M$. Then there exists an $r > 0$ so that the closed ball $\bar{B}(0, r) \subset W_p$ and hence $\exp_p|_{\bar{B}(0,r)}$ is a diffeomorphism on its image.

Before we are ready to do Riemannian geometry properly by introducing a distance function, we can introduce a certain coordinate system that allows us to view things "from a geodesic's point of view". Choose a basis $(E_1, \ldots, E_N)$ of T_pM so that it is a generalized orthonormal basis with respect to $g|_p$, i.e. $g|_p(E_i, E_j) = \delta_{ij}$. Define the linear isomorphism

$$A : \mathbb{R}^n \to T_pM, \qquad (a^1, \ldots, a^n) \mapsto a^i E_i.$$

Definition 3.49. Let $V_p := A^{-1}(W_p) \subset \mathbb{R}^n$ so that $\exp_p \circ A : V_p \to U_p$ is a diffeomorphism with associated chart $x := (\exp_p \circ A)^{-1} : U_p \to V_p$. This means that we have

$$\mathbb{R}^n \supset V_p \overset{A}{\cong} \underset{W_p}{\overset{T_pM}{\cup}} \overset{\exp_p}{\cong} U_p \subset M.$$

Then the coordinate system belonging to x (i.e. its components) are called **Riemannian normal coordinates** or **geodesic normal coordinates** (since the exponential function determines geodesics on W_p).

Why are these coordinates useful for doing geometry?

Proposition 3.50. Let (M, g) be a Riemannian manifold and $p \in M$. Let $g_{ij} : V_p \to \mathbb{R}$ be the matrix representation of the metric g (in $\mathbb{R}^n$ rather than T_pM since these spaces are isomorphic) and let $\Gamma^k_{ij} : V_p \to \mathbb{R}$ be the Christoffel symbols in Riemannian/geodesic normal coordinates around p. Then we (conveniently) have:

$$\text{(i)}\ \ x(p) = 0,$$

$$\text{(ii)}\ \ g_{ij}(0) = \delta_{ij},$$

$$\text{(iii)}\ \ \Gamma^k_{ij}(0) = 0.$$

Proof. (i) Since we have $\exp_p^{-1}(p) = 0$ we find $x(p) = A^{-1}\big(\exp_p^{-1}(p)\big) = A^{-1}(0) = 0$.

(ii) Denote $(e_1, \ldots, e_n)$ the standard basis of $\mathbb{R}^n$. We find

$$\begin{aligned} g_{ij}(0) &= g|_p\big(dx^{-1}|_0(e_i), dx^{-1}|_0(e_j)\big) \\ &= g|_p\big(d(\exp_p \circ A)|_0(e_i), d(\exp_p \circ A)|_0(e_j)\big) \\ &= g|_p\big(d\exp_p|_0(E_i), d\exp_p|_0(E_j)\big) \end{aligned}$$

which, by Lemma 3.46 is equal to $g|_p\big(E_i, E_j\big) = \delta_{ij}$.

(iii) Let $v = (v^1, \ldots, v^n) \in R^n$. Then the geodesic with $c(0) = p$ and $\dot{c}(0) = A(v)$ is $c(t) = x^{-1}(tv) = \exp_p(tA(v))$. In Riemannian/geodesic normal coordinates we get by Example 3.13

$$(\ddot{c})^k + \Gamma^k_{ij}\big(c^1, \ldots, c^n\big) \cdot \dot{c}^i \cdot \dot{c}^j = 0$$

with $c^k = x^k \circ c = tv^k$ so that $\dot{c}^k(t) = v^k$ and $(\ddot{c})^k(t) = 0$. Evaluating in $t = 0$ yields

$$0 = 0 + \Gamma^k_{ij}(0, \ldots, 0) \cdot v^i \cdot v^j.$$

This already hints towards the fact that the Christoffel symbols vanish. We can formally show this by defining another bilinear form h^k on $\mathbb{R}^n$ by $h^k(y, z) = \Gamma^k_{ij}(0) y^i z^j$. To show the symmetry, note that

$$h^k(x, y) = \Gamma^k_{ij}(0) y^i z^j = \Gamma^k_{ji}(0) y^j z^i = \Gamma^k_{ij} y^j z^i = h^k(z, y)$$

where we used the torsion-freeness of the Levi-Civita connection. By the above, we see $h^k(v, v) = 0$ for all $v \in \mathbb{R}^n$ and hence $h^k \equiv 0$ on $\mathbb{R}^n$. This implies that $\Gamma^k_{ij}(0) = 0$ for all $i, j, k = 1, \ldots, n$.

□

Remark. These normal coordinates around p allow us to describe the metric so that it behaves like the standard Euclidean metric. As a result, we can write geodesics pointing radially out of p (*radial geodesics*) as

$$c_p^r(t) = \left(t \frac{\partial}{\partial x^1}\bigg|_p, \ldots, t \frac{\partial}{\partial x^n}\bigg|_p \right).$$

3.5 Riemannian Distance

So far, we have seen that we can define a metric on Riemannian manifolds and that it can be interpreted as the amount of "stretching" or "retracting" with respect to a certain basis. We have seen that there always is an energy-minimizing path (geodesic) between two points on a Riemannian manifold since we can assign to each point a unique Levi-Civita connection. The next step is to determine the *distance* between points on a Riemannian manifold. In this section, we generally assume that (M, g) is a Riemannian manifold.

Definition 3.51. Let $c : [a, b] \to M$ be a continuous and piecewise C^1-curve. Then,

$$L[c] := \int_a^b \|c'(t)\| dt$$

is the **length** of c, where $\|.\|$ is the norm induced by the Riemannian metric, i.e. $\|x\| = \sqrt{g(x, x)}$ and $'$ denotes the derivative with respect to parameter t.

Note that the length of c is independent of its parametrization. If $\psi : [a, b] \to [\alpha, \beta]$ is a parameter transformation, then

$$\begin{aligned} L[c \circ \psi] &= \int_a^b \|(c \circ \psi)'(t)\| dt \\ &= \int_a^b \|c'(\psi(t))\| \cdot |\psi'(t)| dt \\ &\overset{*}{=} \int_\alpha^\beta \|c'(s)\| ds \\ &= L[c] \end{aligned}$$

where $*$ denotes the substitution $s = \psi(t)$. We can define the distance between two points as the length of the shortest curve connecting them, or rather, the infimum of all such curves.

Definition 3.52. Let $p, q \in M$. Then the **Riemannian distance** or **geodesic distance** between p and q is given by

$$d(p, q) := \inf \left\{ L[c] \mid c : [a, b] \to M \text{ a piecewise } C^1\text{-curve with } c(a) = p, c(b) = q \right\}.$$

Note that there need not be a *shortest* curve between two points. Take, for example, $p, q \in M = \mathbb{R}^2 \backslash \{0\}$ so that $p = -q$. Then we have $d(p, q) = 2\|p\|$ but the curve with this length necessarily passes through 0, hence it is not defined on M. Hence all curves c through p and q have $L[c] > d(p, q)$.

Let us show that (M, d) is a metric space.

Lemma 3.53. (Gauss Lemma). Let $p \in M$ and $b > 0$. Assume that the geodesic $c(t) = \exp_p(t\xi)$ is defined on $[0, b]$. Then $\exp_p$ is defined on an open neighborhood of $\{t\xi | 0 \leq t \leq b\} \subset T_pM$ and we have

(i) $d\exp_p |_{t\xi}(\xi) = \dot{c}(0)$,

(ii) for $\eta \in T_{t\xi}(T_pM) \cong T_pM$ we have

$$g\big(d\exp_p |_{t\xi}(\eta), \dot{c}(t)\big) = g(\eta, \xi).$$

In particular, it holds that $d\exp_p |_{t\xi}(\eta) \perp \dot{c}(t)$ if $\eta \perp \xi$.

Proof. (i) We see $d\exp_p |_{t\xi}(\xi) = \frac{d}{ds}\exp_p(t\xi + s\xi)|_{s=0} = \frac{d}{ds}c(t+s)|_{s=0} = \dot{c}(t)$.

(ii) To prove this, one needs the notion of Jacobi fields. We will not treat this here and refer to [2].

□

Collorary 3.54. (without proof). Let $r > 0$ so that $\exp_p |_{\bar{B}(0,r)}$ is a diffeomorphism on its image. Let $c : [a, b] \to M$ be a piecewise C^1-curve that starts in p but ends outside of the image of the (open) ball with radius r around 0 under the exponential function, i.e. $c(a) = p$ and $c(b) \notin \exp_p\big(B(0, r)\big)$. Then we have that $L[c] \geq r$.

The above result seems reasonable: the length of a path starting in p and ending outside of the exponential of the ball with radius r cannot be shorter in length than r.

Theorem 3.55. *(M, d) is a metric space.*

Proof. Obviously, we have $d(p, q) \geq 0$ and $d(p, p) = 0$ (since the constant curve has length 0) for all $p, q \in M$. Let $p \neq q$. Choose $r > 0$ so that the exponential restricted to the close ball around 0 with radius r is a diffeomorphism, and so that $q \notin \exp_p\big(B(0, r)\big)$. By Corollary 3.54 we have that every curve between p and q has at least length r, hence $d(p, q) \geq r > 0$. Thus we have $d(p, q) = 0 \iff p = q$.

Symmetry $d(p, q) = d(q, p)$ follows from the fact that the inverse curve of c, c^-, simply means traversing it in the opposite direction, i.e. $c^-(t) = c(b - t)$ for $c : [a, b] \to M$. The length stays the same.

We are yet to show the triangle inequality. Let $\epsilon > 0$ and choose continuous piecewise C^1-curves c_1 from p to $m \in M$ with $L[c_1] \leq d(p, m) + \epsilon$ and c_2 from m to q with $L[c_2] \leq d(m, q) + \epsilon$. Now concatenate c_1 and c_2 (i.e. consecutively traverse them) so that we have a continuous piecewise C^1 curve c from p to q. We obtain

$$d(p, q) \leq L[c] \leq L[c_1] + L[c_2] \leq d(p, m) + d(m, q) + 2\epsilon.$$

The triangle inequality follows from taking the limit $\epsilon \downarrow 0$. □

Let us now introduce a useful definition.

Definition 3.56. Let $p \in M$. Then the **injectivity radius** of M *at* p is given by largest distance r (actually, the supremum) so that the exponential function at p is a diffeomorphism on the open ball $B(0, r)$, denoted

$$\mathrm{injrad}(p) := \sup\{r > 0 | \exp_p |_{B(0,r)} : B(0, r) \to \exp_p\big(B(0, r)\big) \text{ is a diffeomorphism}\}.$$

Remark. The injectivity radius depends on p and can be very large at some points compared to other points.

We will see that the image of an open ball with a radius smaller than the injectivity radius at p under the exponential function at p is actually an open ball with the same radius around p.

Lemma 3.57. Let $0 < r < \text{injrad}(p)$. Then $\exp_p\big(B(0,r)\big) = B(p,r)$.

Proof. Let $q = \exp_p(\xi)$ with $\|\xi\| < r$. Then the map $t \mapsto \exp_p(t\xi)$ with $t \in [0,1]$ is a curve from p to q with length $\|\xi\| < r$. Thus, $d(p,q) < r$ so that $q \in B(p,r)$.
Conversely, if $q \in B(p,r)$ then there is a curve between q and p with a length smaller than r. Hence, by Corollary 3.54 we have that $q \in \exp_p\big(B(0,r)\big)$. □

As it turns out, the Riemannian distance function d actually induces the original topology on M.

Proposition 3.58. The metric d induces the original topology on M, ensuring that d is indeed a continuous map.

Proof. We have to prove that being open with respect to d (d-open) and being open with respect to the topology on M (open) are equivalent.
Claim: Every d-open set is open.
Let $U \subset M$ be d-open. For every $p \in U$ there must exist a $r(p) > 0$ such that $B(p, r(p)) \subset U$. Let $r(p) < \text{injrad}(p)$ (without loss of generality). Then $B(p, r(p)) = \exp_p\big(B(0, r(p))\big)$ is the image of an open subset of T_pM under a diffeomorphism (hence a homeomorphism). Hence, it is open itself. We now have $U = \bigcup_{p\in M} B(p, r(p))$, i.e. a union of open subsets of M, thus U is open.
Claim: Every open set is d-open.
The proof is similar to the first claim. □

Collorary 3.59. The map $d : M \times M \to \mathbb{R}$ is continuous.

Finally, we will show that every shortest curve between two points is a geodesic (although, as we have seen, the converse need not be true).

Lemma 3.60. Let M be a Riemannian manifold and $c : [a,b] \to M$ a continuous, piecewise C^1-curve. Then

$$L[c]^2 \leq 2(b-a) \cdot E[c]$$

where the equality holds if c is a geodesic.

Proof. This follows from the Cauchy-Schwarz inequality on square-integrable functions (remember that $\|.\| = \sqrt{g(\cdot,\cdot)}$):

$$L[c]^2 = \left(\int_a^b \|\dot{c}\| \cdot 1dt\right)^2 \leq \int_a^b \|\dot{c}\|^2 dt \cdot \int_a^b 1dt = E[c] \cdot 2(b-a).$$

By Cauchy-Schwarz, equality holds if and only if 1 and $\|\dot{c}\|$ are linearly dependent, i.e. if $g(\dot{c},\dot{c})$ is a non-zero constant (positive since g is positive definite). We call c in this case "parametrized to proportionally to arc-length". By Lemma 3.40, geodesics have this property. □

Collorary 3.61. A curve c is a geodesic if and only if c minimizes the length and is parametrized proportionally to arc-length, i.e. $g(\dot{c},\dot{c}) = \lambda$ for some constant $\lambda > 0$.

Collorary 3.62. By the above corollary, if c is the shortest curve connecting p and q, it is a geodesic if and only if it is parametrized proportionally to arc-length.

The converse need not be true. For example, great circles on S^n are geodesics that connect a point to itself. However, there is only one curve with the shortest length that connects a point to itself, and it has length 0 – the constant curve.
In light of the discussion on the heat equation – how the heat of a candle dissipates radially, that is, along geodesics in $\mathbb{R}^3$ – we can generally predict that the heat will spread out along geodesics in any Riemannian manifold. The above tells us that if there is a shortest curve between two points, the heat will travel along that curve.
Let us conclude this chapter with definitions that come in useful with discussing the heat equation and Brownian motion on Riemannian manifolds.

Definition 3.63. A geodesic $c : [a, b] \to M$ is called **minimal** if

$$L[c] = d\big(c(a), c(b)\big).$$

Definition 3.64. Let $c : [0, b] \to M(b > 1)$ be a geodesic so that $p := c(a) \in M$ and so that $c(t) := \exp_p(t\xi)$ for some $\xi \in T_pM$. Then, the set of tangent vectors ξ for which c is a *minimal* geodesic for $t \in [0, 1]$ but *not minimal* for $1 < t < b$ is called the **cutlocus in** T_pM. The **cutlocus in** M is defined as

$$\exp_p\big(\text{cutlocus of } T_pM\big).$$

We may interpret the cutlocus of p in M as the points in the manifold where geodesics starting at p stop being minimizing (hence, it is easier to describe points *outside* of the cutlocus – we will get back to this in Chapter 6).

Definition 3.65. Let $p \in M$. We call M **geodesically complete at** p if $\exp_p$ is defined everywhere on T_pM, that is, if all geodesics through p are defined everywhere on $\mathbb{R}$. If M is geodesically complete at all $q \in M$ we call M **a geodesically complete Riemannian manifold**.

As it turns out, by the Hopf-Rinow Theorem [2] M being a complete Riemannian manifold is equivalent to (M, d) being a complete metric space, and equivalent to M being geodesically complete at a single point. Also by Hopf-Rinow, M is a complete Riemannian manifold if all closed balls of the form $\bar{B}(p, r)$ with $p \in M, r > 0$ are compact.
It follows that any compact Riemannian manifold is geodesically complete.

Part II

Heat Equation and Brownian Motion

4. Heat Equation on Riemannian Manifolds

This chapter contains the discussion on the heat equation on Riemannian manifolds. In particular, we prove that the *fundamental solution* or *heat kernel* of the heat equation, i.e. the solution of the heat equation whose initial condition is a point heat source, exists and is unique on all Riemannian manifolds. In this chapter, we will restrict ourselves to compact Riemannian manifolds without boundary (the case where the compact Riemannian manifold has a boundary will be discussed in Chapter 5). To be able to discuss all this, we need theory on differential- and Riemannian geometry from the preceding chapters.

4.1 The heat equation

The heat equation is a partial differential equation that describes how heat is distributed over space and how that distribution changes over time. Usually, solutions to this equation can be interpreted as temperature. By the second law of thermodynamics, heat will always move from hotter areas to colder areas. Intuitively, the rate at which the temperature changes is proportional to the change in the *slope* of the spatial temperature distribution, i.e. a second spatial derivative. On $\mathbb{R}$, the heat equation reads

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}.$$

More generally, the heat equation on $\mathbb{R}^n$ reads

$$\frac{\partial u}{\partial t} = \Delta u \tag{4.1}$$

where $\Delta := \sum_{i=1}^n \frac{\partial^2}{\partial x^{i2}}$ is the Laplace operator on $\mathbb{R}^n$ representing the second spatial derivative. Given an open subset $U \subset \mathbb{R}^n$ and an open interval $I \subset \mathbb{R}$, a function $u : U \times I \to \mathbb{R}$ is called a *solution of the heat equation* if it is C^1 in the t-parameter, C^2 on the spatial parameter and satisfies Equation (4.1). Oftentimes, the heat equation is described using an operator.

Definition 4.1. The **heat operator** is the operator

$$L := \Delta - \frac{\partial}{\partial t}$$

so that the heat equation (4.1) becomes

$$Lu = 0.$$

Naturally, to describe the flow of heat in a certain body, we must have some initial condition as to where the heat is concentrated at $t = 0$ (else, there is no heat to distribute!). Hence, the (homogeneous) heat equation is given by

$$\begin{cases} Lu(x,t) = 0 & (x,t) \in \mathbb{R}^n \times (0,\infty) \\ u(x,0) = f(x) & x \in \mathbb{R}^n \end{cases} \tag{4.2}$$

for some initial distribution $f(x)$. We aim to study the heat equation on Riemannian manifolds. By Chapters 2 and 3, we know how to differentiate a solution u with respect to t, but the Laplace operator is more subtle. The next section includes theory on the Laplace operator on Riemannian manifolds (M,g) called the *Laplace-Beltrami* operator induced by the metric g, denoted Δ_g. Then, we will dive into the fundamental solution on $\mathbb{R}^n$ before we discuss the fundamental solutions on compact Riemannian manifolds. We largely follow the discussions from [3], [5] and [6].

4.2 The Laplace-Beltrami operator

In this section, we assume that (M,g) is an n-dimensional Riemannian manifold. Let $f \in C_p^\infty(M)$ and remember that the differential at $p \in M$, $df|_p : T_pM \to \mathbb{R}$ is linear, meaning $df|_p \in T_p^*M$. For a chart $x : U \to V$ we have

$$df = \frac{\partial f}{\partial x^i} dx^i.$$

The non-degenerateness of all $g|_p$ implies that there exists a unique vector field on the tangent bundle TM, called the *gradient* of f and denoted by grad f so that

$$g|_p\big((\text{grad } f)(p), \xi\big) = df|_p(\xi) \quad \text{for all } \xi \in T_pM.$$

Let us formally define this.

Definition 4.2. The **gradient** induced by metric g is the linear operator

$$\text{grad } : C_p^\infty(M) \to \Xi_p$$

so that

$$g|_p\big((\text{grad } f)(p), \eta\big) = df|_p(\eta) \quad \text{for all } \eta \in \Xi_p.$$

Let us express the gradient vector field in terms of chart x (i.e. local coordinates). Write, with coefficient functions a^i,

$$\text{grad } f = a^i \frac{\partial}{\partial x^i}.$$

Then we find

$$\begin{aligned} \frac{\partial f}{\partial x^j} &= \frac{\partial f}{\partial x^i} dx^i \left(\frac{\partial}{\partial x^j} \right) \\ &= df \left(\frac{\partial}{\partial x^j} \right) \\ &= g\left(a^i \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right) \\ &= a^i g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right) \\ &= a^i g_{ij}. \end{aligned}$$

Now, we can multiply both sides with the inverse matrix of g_{ij}, namely $(g^{ij})_{ij}$, to find $a^i = g^{ij}\frac{\partial f}{\partial x^j}$. Hence, we obtain

$$\text{grad } f = \frac{\partial f}{\partial x^j} g^{ij} \frac{\partial}{\partial x^i}. \tag{4.3}$$

Next, we discuss the *divergence* operator that acts on vector fields in Ξ_p and returns a function in $C^\infty(M)$.

Definition 4.3. The **divergence** induced by metric g is the linear operator

$$\text{div}_g \ : \Xi_p \to C_p^\infty(M),$$

defined in local coordinates as, for a chart $x : U \to V$ with $p \in U$,

$$\text{div}_g(\eta) = dx^i \left(\nabla_{\frac{\partial}{\partial x^i}} \eta \right)$$

with Levi-Civita Connection ∇ (we leave out evaluations at p here and will do so in the following).

Let us rewrite this expression in terms of the metric g [7]. First we write $\eta = \eta^j \frac{\partial}{\partial x^j}$. By Equation (3.7) we can write

$$\nabla_{\frac{\partial}{\partial x^i}} \eta = \frac{\partial \eta^j}{\partial x^i}\frac{\partial}{\partial x^j} + \eta^j \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \frac{\partial \eta^j}{\partial x^i}\frac{\partial}{\partial x^j} + \eta^j \Gamma^k_{ij} \frac{\partial}{\partial x^j}$$

and thus, since $dx^i(\frac{\partial}{\partial x^j}) = \delta^i_j$, we find

$$\text{div}_g(\eta) = \frac{\partial \eta^i}{\partial x^i} + \eta^j \Gamma^i_{ij}$$

(note how this turns into the regular divergence in $\mathbb{R}^n$, where the Christoffel symbols vanish).

Next, we apply Equation (3.11) to find

$$\begin{aligned} \Gamma^i_{ij} &= \frac{1}{2} g^{ik} \left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k} \right) \\ &= \frac{1}{2} g^{ik} \frac{\partial g_{ik}}{\partial x^j} \\ &= \frac{1}{2} \text{Tr}\left((g^{ik}) \frac{\partial}{\partial x^j} (g_{ik}) \right) \\ &\overset{*}{=} \frac{1}{2} \frac{\frac{\partial}{\partial x^j} \det(g_{ik})}{\det(g_{ik})} \end{aligned}$$

where, at $*$, we used the theorem from [8] and the fact that g^{ik} is the inverse of g_{ik}. Hence, by simply rewriting conveniently and using the product rule backwards, we arrive at

$$\begin{aligned} \text{div}_g(\eta) &= \frac{\partial \eta^i}{\partial x^i} + \eta^j \Gamma^i_{ij} \\ &= \frac{1}{\sqrt{\det(g_{kl})}} \sqrt{\det(g_{kl})} \frac{\partial \eta^i}{\partial x^i} + \frac{1}{\sqrt{\det(g_{kl})}} \frac{1}{2\sqrt{\det(g_{kl})}} \frac{\partial}{\partial x^j} \big(\det(g_{kl}) \eta^j \big) \\ &= \frac{1}{\sqrt{\det(g_{kl})}} \frac{\partial}{\partial x^j} \left(\eta^j \sqrt{\det(g_{kl})} \right). \end{aligned} \tag{4.4}$$

We are now equipped to define the Laplace-Beltrami operator.

Definition 4.4. The **Laplace-Beltrami** operator on (M, g) is the linear operator

$$\Delta_g : C_p^\infty(M) \to C_p^\infty(M)$$

defined by

$$\Delta_g := \mathrm{div}_g \circ \mathrm{grad}\ .$$

Its linearity follows from the linearity of div_g and grad. Moreover, we have that

$$\Delta_g(f \cdot h) = f \cdot \Delta_g(h) + \Delta_g(f) \cdot h + 2g(\mathrm{grad}\ f, \mathrm{grad}\ h). \tag{4.5}$$

We express Δ_g in terms of chart x as, using Equations (4.3) and (4.4),

$$\Delta_g = \frac{1}{\sqrt{\det(g_{kl})}} \frac{\partial}{\partial x^i} \left(g^{ij} \sqrt{\det(g_{kl})} \frac{\partial}{\partial x^j} \right). \tag{4.6}$$

We can also look at the Laplace-Beltrami operator in geodesic normal coordinates. Fix $p \in M$. Denote the associated chart with y. By Proposition 3.50 and using Lemma 3.57 we find that $g_{ij} = \delta_{ij}$ with determinant 1 for all $q \in \mathrm{injrad}(p)$ so that its derivatives with respect to the components of y vanish. Hence, we find

$$\Delta_g f = \sum_{i=1}^{n} \frac{\partial f}{\partial y^i}$$

in normal geodesic coordinates.

Let us look at some examples.

Example 4.5. Consider $M = \mathbb{R}^n$. By example 3.11 we have that $(g_{\text{Eucl}})_{ij} = \delta_{ij}$ for all $x \in \mathbb{R}^n$. Plugging this in Equation (4.6) we obtain straightforwardly

$$\Delta_{g_{\text{Eucl}}} = \sum_{i=1}^{n} \frac{\partial^2}{\partial x^{i2}}$$

as we have seen in Section 4.1.

Example 4.6. Consider $M = S^2$, i.e. the 2-sphere. Consider the standard metric in spherical coordinates

$$g_{\text{std}}^{\text{sph}}(\phi, \theta) = \begin{pmatrix} \sin^2\theta & 0 \\ 0 & 1 \end{pmatrix}$$

from Example 3.16. Its determinant is equal to $\sin\theta$. Hence, we get by Equation (4.6)

$$\begin{aligned} \Delta_{g_{\text{std}}^{\text{sph}}} &= \frac{1}{\sin\theta} \left(\frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} \right) + \frac{\partial}{\partial\phi} \left(\frac{1}{\sin\theta} \frac{\partial}{\partial\phi} \right) \right) \\ &= \frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\phi^2}. \end{aligned} \tag{4.7}$$

4.3 Heat equation on Riemannian manifolds and the Fundamental Solution

Let (M, g) be a compact Riemannian manifold without boundary throughout the rest of this chapter. We define the heat operator for Riemannian manifolds similar to the definition on $\mathbb{R}^n$.

Definition 4.7. The **heat operator** on (M, g) is the operator

$$L := \Delta_g - \frac{\partial}{\partial t}$$

so that the heat equation becomes

$$\begin{cases} Lu(x,t) = 0 & (x,t) \in M \times (0, \infty) \\ u(x,0) = f(x) & x \in M \end{cases} \tag{4.8}$$

for some initial distribution $f(x)$.

Before we go ahead and prove some interesting properties of the Riemannian manifold version of the heat equation, we must define how to integrate over (subsets of) a Riemannian manifold. The exact definition involves *differential forms* which we will not go into here.

Definition 4.8. Let J be a finite indexing set so that $\{U_j, x_j\}_{j \in J}$ is an atlas of M (this is possible since M is assumed to be compact). Moreover, let $\{\rho_j\}_{j \in J}$ be a partition of unity subordinate to $\{U_j, x_j\}_{j \in J}$. We can represent an (infinitesimal) **volume element** of (M, g) by the differential form ω_g. Therefore, we define the **integral over** M as

$$\int_M \omega_g = \sum_{j \in J} \int_{U_j} \rho_j \omega_g := \sum_{j \in J} \int_{x_j(U_j)} \left(x_j(\rho_j \omega_g) \right)$$

where we can interpret the latter integral as an integral over a subset of $\mathbb{R}^n$ for infinitesimal volume elements $dV_j := x_j(\rho_j \omega_g)$ in the usual sense. In short, we use the compactness and the homeomorphisms with subsets of $\mathbb{R}^n$ to formalize an integration technique. Instead of writing the ω_g differential form, we will write a volume element (depending on spatial variable x and metric g) as $dV_g(x)$ (cf. [5]).

First, let us show that the solution to the heat equation (4.8) is unique. To this end, note that u must be a square-integrable function. Then, we can show that $u(x,t)$ is a decreasing function in t (with respect to the L^2-norm, defined on a manifold M by $\|u\|_{L^2} = \int_M |u(x,t)|^2 dV_g(x)$).

Lemma 4.9. Let $u(x,t)$ solve the heat equation. Then $t \mapsto \|u(\cdot,t)\|_{L^2}$ is decreasing with t.

Proof. We have that

$$\begin{aligned} \frac{d}{dt} \|u(\cdot,t)\|_{L^2}^2 &= \int_M \frac{\partial}{\partial t} \big(u(x,t) u(x,t) \big) dV_g(x) \\ &= 2 \int_M u(x,t) \frac{\partial}{\partial t} u(x,t) dV_g(x) \\ &= -2 \int_M u(x,t) \Delta_g u(x,t) dV_g(x) \\ &= -2 \|\Delta_g u(\cdot,t)\|_{L^2} \leq 0. \end{aligned}$$

□

Lemma 4.10. Let $u(x,t)$ solve the heat equation (4.8). Then the solution is unique.

Proof. Suppose that u_1 and u_2 are two are solutions, then $u := u_1 - u_2$ is a solution to

$$\begin{cases} Lu(x,t) = 0 & (x,t) \in M \times (0,\infty) \\ u(x,0) = 0 & x \in M \end{cases}.$$

Since $u(x,t)$ is a decreasing function in t and $\|u(\cdot,0)\|_{L^2} = 0$, we must have that $\|u(\cdot,t)\|_{L^2} = 0$ for all $t > 0$ and hence $u(x,t) = 0$ for all $x \in M, t \in (0,\infty)$. Hence $u_1 = u_2$. □

Next, we introduce Duhamel's Principle. This is a mathematical trick that we use to show uniqueness of the fundamental solution (which will be formally defined later on).

Proposition 4.11. (Duhamel's Principle). Let $u, v : M \times (0,\infty) \to \mathbb{R}$ be maps that are C^2 in the first parameter and C^1 on the second. Let $t > 0$ and let a, b such that $a < b$ and $[a,b] \subset (0,t)$. Then, we have

$$\int_M u(y,t-a)v(y,a) - u(y,t-b)v(y,b)dV_g(y)$$
$$= \int_a^b \int_M Lu(y,t-s)v(y,s) - u(y,t-s)Lv(y,s)dV_g(y)\, ds.$$

Proof. We write out

$$\begin{aligned} &Lu(y,t-s)v(y,s) - u(y,t-s)Lv(y,s) \\ &\quad = \Delta_g u(y,t-s)v(y,s) - u(y,t-s)\Delta_g v(y,s) + \frac{\partial}{\partial s}u(y,t-s)v(y,s) - u(y,t-s)\frac{\partial}{\partial s}v(y,s) \\ &\quad = \Delta_g u(y,t-s)v(y,s) - u(y,t-s)\Delta_g v(y,s) - \frac{\partial}{\partial s}\big(u(y,t-s)v(y,s)\big). \end{aligned}$$

First of all, upon integrating over the spatial parameter y we find that the first two terms cancel each other out. This follows from the fact that the Laplace-Beltrami operator is *formally self-adjoint*. This property is a direct result of Green's theorem and the fact that, in this case, the boundary of M is empty. For a detailed derivation of formal self-adjointness, see [6].
Next, integrating with respect to s from a to b gives

$$\left(\int_M \big(u(y,t-s)v(y,s)\big) \right) \Bigg|_{s=a}^{s=b}$$

which yields the wanted result. □

Let us now define the fundamental solution of the heat equation. Assuming, for now, that such a solution exists, we will show that it must be unique. Existence will take more rigor to prove.

Definition 4.12. Let L_x be the heat operator where the Laplace-Beltrami operator acts on the variable x. The **fundamental solution** (also called **heat kernel**) is given by a continuous function $p : M \times M \times (0, \infty) \to \mathbb{R}$ which is C^2 with respect to spatial parameters $x, y \in M$ and C^1 with respect to the temporal parameter t and satisfies

$$L_x p(x, y, t) = 0 \quad \text{and} \quad \lim_{t \to 0} p(\cdot, y, t) = \delta_y$$

where δ_y is the Dirac delta function at $y \in M$ i.e. it is infinite at y and zero elsewhere and has integral 1, so that we have

$$\lim_{t \to 0} \int_M p(x, y, t) f(x) dV_g(x) = f(y)$$

for all bounded maps $f : M \to \mathbb{R}$. We can interpret the fundamental solution as a solution to the heat equation that describes how heat dissipates from a *point source* at $y \in M$ with a temperature of 1. Then, for any initial distribution $f(y)$ we can find a solution by "superposition": we think of the $f(y)$ as an infinite set of infinitesimal point sources with finite temperature (≤ 1) and integrate over all of them:

$$u(x, t) = \int_M p(x, y, t) f(y) dV_g(y).$$

Proposition 4.13. The fundamental solution to the heat equation (4.8) on M is symmetric in the two spatial variables x, y. Moreover, it is unique (here, we assume that it exists).

Proof. Suppose that there are two fundamental solutions p_1 and p_2. Fix the points $x, z \in M$. We are going to apply Duhamel's principle to $u(y, t) = p_1(x, y, t)$ and $v_(y, t) = p_2(z, y, t)$. Note that the right hand side of Duhamel's principle vanishes since $Lp_1 = Lp_2 = 0$. Thus, we have

$$\int_M p_1(x, y, t - a) p_2(z, y, a) - p_1(x, y, t - b) p_2(z, y, b) dV_g(y) = 0.$$

Now, let the parameters a and b approach 0 and t, respectively. Then $p_2(z, y, a) \to \delta_x$ and $p_1(x, y, t - b) \to \delta_z$, hence we get that $p_1(x, z, t) - p_2(z, x, t) = 0$ or

$$p_1(x, z, t) = p_2(z, x, t).$$

We can repeat the previous steps with a single fundamental solution p (i.e. set $p_1 = p = p_2$) to find that for all fundamental solutions p we have $p(x, z, t) = p(z, x, t)$, that is, they are symmetric in the spatial variables. Now we know that $p_1(x, z, t) = p_1(z, x, t) = p_2(z, x, t)$ and therefore we have that p_1 must agree with p_2 on $M \times M \times (0, \infty)$. Thus, the fundamental solution is unique. □

Collorary 4.14. Since the fundamental solution $p(x, y, t)$ is symmetric in the spatial variables x and y, the operators L_x and L_y act equivalently on p.

An interesting property of the fundamental solution is that it integrates (spatially) to 1. To show this, we introduce the *inhomogeneous* heat equation and we prove a useful lemma.

Definition 4.15. Let $f : M \to \mathbb{R}$ be C^2 and let $F : M \times (0, \infty)$ be C^2 on M and C^1 on $(0, \infty)$ (assume both are not trivial maps). Then the **inhomogeneous heat equation** is given by

$$\begin{cases} Lu(x, t) = F(x, t) & (x, t) \in M \times (0, \infty) \\ u(x, 0) = f(x) & x \in M \end{cases}. \tag{4.9}$$

We also refer to Equation (4.8) as the **homogeneous** heat equation.

Remark notice that the inhomogeneous heat equation also has a unique solution. Just repeat the proof of the homogeneous case ($L(u_1 - u_2)$ must then be zero as well).

Physically, we can interpret $Lu(x,t) = F(x,t)$ as the presence of some kind of external heating or refrigeration process governed by F.

Lemma 4.16. Let $f : M \to \mathbb{R}$ be C^2 and let $F : M \times (0,\infty)$ be C^2 on M and C^1 on $(0,\infty)$. Then the solution to inhomogeneous heat equation (4.9) is given by

$$u(x,t) := \int_M p(x,y,t)f(y)dV_g(y) + \int_0^t \int_M p(x,y,s)F(y,t-s)dV_g(y)\,ds.$$

Proof. Let $x \in M$ be fixed and let u be the solution to the inhomogeneous heat equation (4.9). We apply Duhamel's principle to u and $v(y,t) := p(x,y,t)$. Then, since $Lu = F$ and $Lp = 0$ we obtain

$$\int_M u(y,t-a)p(x,y,a) - u(y,t-b)p(x,y,b)dV_g(y) = \int_a^b \int_M F(y,t-s)p(x,y,s)dV_g(y)\,ds.$$

Now, let $a \to 0$ and $b \to t$ so that the first term of the integrand in the left hand side vanishes and $u(y,t-b) \to f(y)$. Then rearranging the equation yields the desired expression. □

Collorary 4.17. For a fundamental solution p we have

$$\int_M p(x,y,t)dV_g(y) = 1.$$

Proof. Let $u(x,t) \equiv 1$ so that it satisfies the homogeneous heat equation (i.e. $F \equiv 0$) with initial distribution $f \equiv 1$. Then, by Lemma 4.16 we get the assertion immediately. □

Collorary 4.18. Let $t, s > 0$. Then, for fundamental solution p we have

$$p(x,z,t+s) = \int_M p(x,y,t)p(y,z,s)dV_g(y).$$

Proof. Let $y \in M$ be fixed. Then $u(z,s) := p(z,y,t+s)$ solves the homogeneous heat equation with initial distribution $f(x) = p(x,y,t)$. Hence, by Lemma 4.16 we get the assertion. □

4.4 The fundamental solution on $\mathbb{R}^n$

The reader might be confused by the fact that we have been talking about the heat equation on Riemannian manifolds and then "suddenly" turn to $\mathbb{R}^n$ again. The reason for this is that we give the process of arriving at the existence (and formulas) of the fundamental solutions for compact Riemannian manifolds rather than just state it and prove it directly. The fundamental solution on $\mathbb{R}^n$ is not only a good warm-up, it provides a very useful ansatz for determining the fundamental solution on compact Riemannian manifolds.
Before continuing, let us give a brief reminder on basic Fourier transform theory.

Definition 4.19. Let $f \in C^{\infty}(\mathbb{R}^n)$, then its **Fourier Transform** is given by

$$\hat{f}(\xi) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-ix\xi} f(x) dx.$$

Remark. In the above definition, f need not be a function, but can be a distribution as well. For example, the Fourier transform of 1 is a Dirac delta function.

Definition 4.20. Let $f, g \in L^2(\mathbb{R}^n)$, i.e. square-integrable functions. Then their **convolution product** is given by

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

Lemma 4.21. Some well-known results on Fourier Transforms include:

(i) $(f * g)\hat{} = \hat{f} \cdot \hat{g}$

(ii) $(f \cdot g)\hat{} = \hat{f} * \hat{g}$

(iii) $f(x) = \int_{\mathbb{R}^n} e^{ix\xi} \hat{f}(\xi) d\xi$

(iv) $(f')\hat{}(x) = i \cdot \|x\| \hat{f}(x)$.

Assume now that $p(x, y, t)$ is the fundamental solution on $\mathbb{R}^n$. Then, for initial distribution $f(x)$,

$$u(x,t) = \int_{\mathbb{R}^n} p(x, y, t) f(y) dy$$

solves homogeneous heat equation (4.2). In the following, $'\hat{\ }'$ will denote a Fourier transform in the spatial variable, $\|.\|$ will denote the Euclidean norm and ∂t will denote the partial derivative with respect to t. Then, we get

$$\begin{aligned} \|y^2\| \hat{u}(y,t) &= \sum_{i=1}^{n} y_i^2 \hat{u}(y,t) \\ &\overset{\text{(iv)}}{=} (-\sum_{i=1}^{n} \frac{\partial^2}{\partial x^{i2}} u)\hat{}(y,t) \\ &= (-\Delta u)\hat{}(y,t) \\ &= -\widehat{\partial_t u}(y,t) = -\partial_t \hat{u}(y,t). \end{aligned}$$

For fixed x and y, this is an ordinary differential equation in the variable t that has the solution

$$\hat{u}(y,t) = h(y) e^{-t\|y\|^2}$$

for some map $h : \mathbb{R}^n \to \mathbb{R}$ so that $h(y) = \hat{u}(y, 0) = \hat{f}(y)$. Hence we write $\hat{u}(y,t) = \hat{f}(y) e^{-t\|y\|^2} = \hat{f}(y) e^{-\|\sqrt{2t}y\|^2/2}$ and we apply (iv) again to find

$$\hat{u}(y,t) = \hat{f}(y) \widehat{\Big(\frac{1}{(2t)^{n/2}} e^{-\|\cdot\|^2/4t} \Big)} (y).$$

Hence, by (i),

$$\hat{u}(\cdot, t) = \widehat{\left(f * \left[\frac{1}{(2t)^{n/2}} e^{-\|\cdot\|^2/4t} \right] \right)}.$$

In other words, we have by definition of the Fourier transform

$$u(y,t) = f * \left[\frac{1}{(2t)^{n/2}} e^{-\|\cdot\|^2/4t}\right](y) = \frac{1}{(4\pi t)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-\|y-x\|^2/4t} dx$$

and therefore we must have

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

and there we have it – the fundamental solution of the heat equation in $\mathbb{R}^n$. We still have to prove this, but first, let's discuss this result. First, observe that $p(x,y,t) \geq 0$ for all $x, y \in \mathbb{R}^n, t \in (0,\infty)$.
Earlier, we motivated that the angle at which you hold your hand near a candle is irrelevant in terms of the temperature you feel, as long as the distance to the candle is constant. However, it seems logical that the temperature you feel *does* depend on the distance. That is, if your hand is only a millimeter away from the fire, it will burn. On the other hand, if your friend Rik who lives 10 kilometres away lights a candle, you cannot feel it. The fundamental solution provides these insights as well - if $x \approx y$, the felt temperature is about 1 (i.e. $u(x,t) = 1$ with $f(y) = \delta_y$). On the other hand, if $\|x\| \gg \|y\|$, the exponential factor ensures that $u(x,t) \to 0$. Now, we can ask ourselves the question: how does the felt temperature change for intermediate distance from the candle (i.e. not too close and not too far) (in this case, we assume that the candle keeps on burning, so that the heat distribution does not change and we can "fix" t)? Notice that p is exactly given by an n-dimensional Gaussian distribution for a fixed t. It is a well-known result (from i.e. statistical physics) that the Gaussian distribution above approaches the Dirac delta function $\delta(y-x) =: \delta_x(y)$ as $t \to 0$. Hence we have the following: the moment a candle is lit ($t = 0$), we have the "point source" of heat in the form of a Dirac delta. After that, the heat spreads out like a Gaussian distribution until it finds a steady-state distribution. When the candle is blown off (somehow without moving the surrounding air), the steady-state distribution slowly approaches a uniform distribution of a temperature of (nearly) 0. The Gaussian nature of the fundamental solution already hints towards probabilistic phenomena such as Brownian motion, but more on that later.

Proposition 4.22. The function $p : \mathbb{R}^n \times \mathbb{R}^n \times (0,\infty) \to [0,\infty)$ given by

$$p(x,y,t) = \frac{1}{(4\pi t)^{n/2}} e^{-\|y-x\|^2/4t} \tag{4.10}$$

is the fundamental solution for the heat equation on $\mathbb{R}^n$.

Proof. It is easy to check that $L_y p = 0$. We have yet to prove that $\lim_{t\to 0} p(x,y,t) = \delta_x(y)$. First, let us show that p integrates to unity. This follows directly from the fact that it is a Gaussian probability distribution, but we show it here nonetheless. We will use integration with "n-spherical coordinates", i.e. we integrate p over a sphere S^{n-1} with radius r and midpoint x and then integrate over all values of r. We denote the "direction" (i.e. all angles that define a point on a sphere) with ξ and we look at $y = x + r\xi$. The Jacobian is known the be r^{n-1} by going from $\int_{\mathbb{R}^n}$ to $\int_0^\infty \int_{S^{n-1}(x)}$, hence

$$\begin{aligned}\int_{\mathbb{R}^n} p(x,y,t)dy &= \int_0^\infty \int_{S^{n-1}(x)} p(x,x+r\xi,t)r^{n-1}d\xi dr \\ &= \int_0^\infty \int_{S^{n-1}(x)} \frac{1}{(4\pi t)^{n/2}} e^{-r^2/4t} r^{n-1} d\xi dr \\ &\overset{*}{=} \int_0^\infty \int_{S^{n-1}(x)} \frac{1}{(4\pi t)^{n/2}} e^{-s^2} s^{n-1}(4t)^{(n-1)/2}\sqrt{4t} d\xi dr \\ &= \pi^{-n/2}\int_0^\infty \int_{S^{n-1}(x)} e^{-s^2}s^{n-1}d\xi ds \\ &= \pi^{-n/2}\int_{\mathbb{R}^n} e^{-\|y\|^2}dy = 1.\end{aligned}$$

where at $*$ we made the substitution $s^2 := r^2/4t$ so that $ds = 2\sqrt{t}dr$ and $r^{n-1} = s^{n-1}(4t)^{(n-1)/2}$. In the last step we used the Gaussian integral and in the step before that we returned to usual integration hence dividing by Jacobian s^{n-1} and introducing the norm.
Now let $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded map. Let $R > 0$ and let $B_\rho(x)$ be the open ball around x with radius $\rho := 2\sqrt{t}R$. Then we have by the triangle inequality

$$\begin{aligned}&\left|f(x) - \int_{\mathbb{R}^n} p(x,y,t)f(y)dy\right| = \left|\int_{\mathbb{R}^n} p(x,y,t)(f(x)-f(y))dy\right| \\ &\le \left|\int_{B_\rho(x)} p(x,y,t)(f(x)-f(y))dy\right| + \left|\int_{\mathbb{R}^n\setminus B_\rho(x)} p(x,y,t)(f(x)-f(y))dy\right| \\ &\le 1\cdot \sup_{y\in B_\rho(x)}|f(x)-f(y)| + \int_\rho^\infty\int_{S^{n-1}} p(x,x+r\xi,t)|(f(x)-f(x+r\xi))|r^{n-1}d\xi dr\end{aligned}$$

where, in the last step, we used the following: in the first term, the integral cannot be larger than the supremum of the integrand multiplied by the integral over p, which is 1. In the second term we switched to n-spherical coordinates and excluded the ball $B_\rho(x)$. Moreover, we used the "triangle inequality" for integrals, i.e. $|\int f(x)dx| \le \int |f(x)|dx$. We are now to show that the expression above vanishes as $R \to \infty$ and $t \to 0$, since then we have that

$$f(x) = \lim_{t\to 0}\int_{\mathbb{R}^n} p(x,y,t)f(y)dy$$

and hence that $\lim_{t\to 0} p(x,y,t) = \delta_x(y)$.
The idea is as follows: we choose R to be very large, but not approaching infinity. Then, we let $t \to 0$ so that $\rho \to 0$ and the first term vanishes. Next, we do the substitution $r \to 2\sqrt{t}s$ (so that we integrate from R instead of ρ). Moreover, if f is bounded by constant C, then $|f(x) - f(x+r\xi)| \le 2C \max_{y\in\mathbb{R}^n\ B_\rho(x)} =: 2C\|f\|_\infty$. Thus, call the second term $T2$, then we have

$$\begin{aligned}T2 &\le \int_R^\infty\int_{S^{n-1}} p(x,x+2\sqrt{t}s,t)\cdot 2\|f\|_\infty \cdot (2\sqrt{t}s)^{n-1}\cdot 2\sqrt{t}d\xi ds \\ &= \int_R^\infty e^{-s^2}s^{n-1}\cdot 2\|f\|_\infty ds \int_{S^{n-1}} d\xi\end{aligned}$$

where we see that the influence of t vanishes by taking into account the Jacobian of the substitution. The integral over the sphere will yield an n-dimensional "volume" that is finite and hence, as $R \to \infty$, this expression vanishes as well. $\square$

Now we are done! Let us get back to the fundamental solution on compact Riemannian manifolds and see how the result in $\mathbb{R}^n$ can help us.

4.5 Fundamental solution on compact Riemannian manifolds

Before diving into the fundamental solution, let us make a setting.
Let $p \in M$ and let $\epsilon = \inf_{p \in M}(\text{injrad}(p))$ so that the exponential map $\exp_p$ is a diffeomorphism on $B_\epsilon(0) \in T_pM$ for all $p \in M$. Then, let $q \in B_\epsilon(p) := \exp_p(B_\epsilon(0))$. We are going to write points in polar coordinates. That is, $q = \exp_p(r\xi)$ with $r \in (0, \epsilon)$ and $\xi \in S^{n-1}(0) \subset T_pM$. We can write the Laplace-Beltrami operator on $\mathbb{R}^n$ from Equation (4.6) can be written in terms of *geodesic* polar coordinates, i.e. a radial operator $\partial_r := \frac{\partial}{\partial r}$ and a Laplace-Beltrami operator for the metric induced on the geodesic sphere with radius r around p, $S_r^{n-1}(p) \subset M$, here denoted g_S, as follows [6]:

$$\Delta_g = \partial_r^2 + \frac{\partial_r\left(\sqrt{\det(g_{ij})}\partial_r\right)}{\sqrt{\det(g_{ij})}} + \Delta_{g_S}.$$

Now, we wish to find the fundamental solution on compact Riemannian manifolds. We are inspired indeed by the fundamental solution on $\mathbb{R}^n$ given by Equation (4.10). Since we have developed a notion of distance on Riemannian manifolds, we can try the exact same solution but in terms of Riemannian distance $d(p,q)$ rather than $\|y - x\|^2$ – the distance on $\mathbb{R}^n$. Note that for a $q = \exp_p(r\xi)$ we have that $d(p,q) = r$ is the length of the geodesic connecting p and q (the "radial" geodesic). Outside of $B_\epsilon(p)$ things become a bit messy since we do not know what the geodesic is and distance becomes harder to define. Therefore, consider the set of all ϵ-balls around all points in M $V_\epsilon := \{(p,q) \in M \times M \mid q \in B_\epsilon(p)\}$. Now, we can imitate the fundamental solution in $\mathbb{R}^n$ (and thus turning back to x, y instead of p, q) by defining the function $G : V_\epsilon \times (0, \infty) \to \mathbb{R}$ defined by

$$G(x,y,t) := \frac{1}{(4\pi t)^{n/2}} e^{-d(x,y)^2/4t}.$$

Now, this function may not be defined everywhere on M but let us ignore that, since we have a bigger problem: G does not solve the heat equation! Let us check this. In geodesic polar coordinates, we have

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

and letting Δ_g work on the variable y we get

$$\Delta_g G = \partial_r^2 G + \frac{\partial_r\left(\sqrt{\det(g_{ij})}\partial_r\right)G}{\sqrt{\det(g_{ij})}} + \Delta_{g_S} = \partial_r^2 G + \partial_r G \frac{\partial_r \sqrt{\det(g_{ij})}}{\sqrt{\det(g_{ij})}}$$

since G is independent of ξ so $\Delta_{g_S} G = 0$.
Next, let us define the auxiliary function $F_x : B_\epsilon(x) \to \mathbb{R}$ by

$$F_x(y) := \frac{\sqrt{\det(g|_y)_{ij}}}{r^{n-1}}, \quad F_x(x) = 1$$

where $g|_y$ is the metric at y with respect to the geodesic polar coordinates coming from x.

Lemma 4.23. The function $F_x : B_\epsilon(x) \to \mathbb{R}$ is smooth and strictly positive.

Proof. This follows from the fact that the metric is positive definite and smooth. By positive-definiteness, we see that the determinant of the metric at y is positive. Moreover, on $B_\epsilon(x)\backslash\{x\}$ we have that $0 < r = d(x,y) < \epsilon$ and hence the function $y \mapsto 1/d(x,y)^{n-1} = 1/r^{n-1}$ is positive and smooth. Hence, so must be F_x. Finally, we are left to show that F_x is continuous in x. If we let $y \to x$, or equivalently $r \to 0$, we see that the metric in normal geodesic coordinates approaches δ_{ij} (Proposition 3.50(ii)). Hence, the metric in geodesic polar coordinates approaches the Euclidean polar metric

$$g^{\text{polar}} = \begin{pmatrix} 1 & & & 0 \\ & r^2 & & \\ & & \ddots & \\ 0 & & & r^2 \end{pmatrix}.$$

The case $n = 2$ was derived in Example 3.11. Note that $\det(g^{\text{polar}}) = r^{2(n-1)}$. Hence, we see that

$$\lim_{y \to x} F_x(y) = \frac{r^{n-1}}{r^{n-1}} = 1$$

and hence F_x is smooth on all of $B_\epsilon(x)$. We conclude in particular that the root, the derivative and the inverse of F_x are well-defined on $B_\epsilon(x)$. □

We now compute

$$\partial_t G = \left(\frac{n}{2t} + \frac{r^2}{4t^2}\right) G$$

while

$$\Delta_g G = -\partial_r^2 G - \partial_r G \left(\frac{\partial_r F_x}{F_x} + \frac{n+1}{r}\right) = \left(\frac{n}{2t} + \frac{r^2}{4t^2}\right) G + \frac{r}{2t}\frac{\partial_r F_x}{F_x} G \neq \partial_t G.$$

Thus, G does not work. Hence, we have two options at this point: start over and try a new solution, or modify G so that it *does* solve the heat equation. The reader may have guessed that we tend to go for the latter option.
Now, how do we go about modifying G? Well, the second term of $\Delta_g G$ makes it so that it does not solve the heat equation. This term will blow up as $t \to 0$. To this end, we can try to multiply G by some function and a power of t. But which power do we choose? Answer: well, all of them! Let, for all $k \in \mathbb{N}$ the function $S_k : V_\epsilon \times (0,\infty) \to \mathbb{R}$ be given by

$$S_k(x,y,t) := G(x,y,t) \sum_{i=0}^{k} u_i(x,y) t^i$$

with maps $u_i : V_\epsilon \to M$ (note that t^i represents "to the power of i"). We hope to determine the u_i so that $L_y S_k = 0$. As mentioned, the spoiler is that we will not get this, but we will make it so that $L_y S_k \to 0$ as $t \to 0$ with "speed of convergence" similar to $t^{k-n/2}$.
Let us compute $L_y S_k$. First, we compute the Laplace-Beltrami operator acting on S_k

using Equation (4.5):

$$\Delta_g(S_k) = \Delta_g G \cdot \Big(\sum_{i=0}^{k} u_i(x,y)t^i\Big) + G \cdot \Big(\sum_{i=0}^{k} \Delta_g(u_i(x,y)t^i)\Big) + 2g\big(\text{grad } G, \text{grad } (u_0 + \ldots + t^k u_k)\big).$$

Since G is independent of ξ, the gradient of G involves only radial derivatives. Also, in geodesic polar coordinates, $g(\partial_r, \partial_\xi) = 0$ and $g(\partial_r, \partial_r) = 1$. (compare to the polar metric in $\mathbb{R}^2$ and the paraboloid in Examples 3.11, 3.15). Hence we get that the last term is equal to $2\partial_r G(\partial_r u_0 + \ldots + t^k \partial_r u_k) = \frac{r}{t}(\partial_r u_0 + \ldots + t^k \partial_r u_k)G$. By the discussion on G not solving the heat equation, it follows that $L_y G = \frac{r}{2t}\frac{\partial_r F_x}{F_x} G$. Hence, we find

$$L_y S_k = G \cdot \Big(\sum_{i=1}^{k} u_i t^{i-1} + \frac{r}{2t}\frac{\partial_r F_x}{F_x} \cdot (\sum_{i=0}^{k} u_i t^i) + \frac{r}{t}(\sum_{i=0}^{k} t^i \partial_r u_i) + \sum_{i=0}^{k} \Delta_g u_i\Big) \tag{4.11}$$

where the first term in the parentheses denotes $\partial_t \sum_{i=0}^{k} u_i t^i$.

What we want to do is build the u_i so that

$$L_y S_k = G \cdot t^k \cdot \Delta_g u_k \tag{4.12}$$

so that it vanishes like $t^{k-n/2}$ (since $G \propto t^{-n/2}$). We can rearrange the terms in Equation (4.11) and "force" it to be equal to Equation (4.12) to find the system of equations

$$r\partial_r u_0 + \frac{r}{2}\frac{\partial_r F_x}{F_x} u_0 = 0$$

$$r\partial_r u_i + \frac{r}{2}\Big(\frac{\partial_r F_x}{F_x} + i\Big) u_i + \Delta_g(u_{i-1}) = 0, \quad i = 1, \ldots, k.$$

The first equation yields $u_0(x,y) = f(\xi) F_x^{-1/2}(y)$ for some function f of the angular variables ξ in geodesic polar coordinates coming from x. However, we want u_0 to be defined on $y = x$ (i.e. $r = 0$) thus f must be constant (independent of the angles). We set it to 1 so that

$$u_0(x,y) = F_x^{-1/2}(y) = \frac{r^{(n-1)/2}}{\big(\det(g|_y)_{ij}\big)^{1/4}}, \qquad u_0(x,x) = 1.$$

Next, we show the existence of u_i for all other $1 \le i \le k$. The second equation in the system is rather hard to solve, thus we use the fact that $v_i = f r^{-i} F_x^{-1/2}$ for a function of angles $f = f(\xi)$ solves the simpler equation

$$r\partial_r v_i + \Big(\frac{r}{2}\frac{\partial F_x}{F_x} + i\Big) v_i = 0.$$

We can then make an "ansatz" by using the chain rule backwards, given by

$$u_i := h r^{-i} F_x^{-1/2}$$

for some function h dependent on distance r, $h = h(r)$. Plugging this into the second equation, we see that u_i satisfies it if we have

$$\partial_r h + F_x^{1/2} \Delta_g (u_{i-1}) r^{i-1} = 0.$$

To wrap up, denote the geodesic between x and y as c. For fixed x, the function $\Delta_g u_{i-1}(x, y)$ is a function of $r = d(x, y)$ along c, hence we can simply find h by integrating with respect to r:

$$h(r) = \int_0^r F_x^{1/2}\big(c(s)\big) \cdot \Delta_g\big(u_{i-1}(c(s), y)\big) s^{i-1} ds$$

and we can conclude that Equation (4.12) is satisfied as long as

$$u_i(x, y) = (d(x, y))^{-i} F_x^{-1/2}(y) \int_0^{d(x,y)} F_x^{1/2}\big(c(s)\big) \cdot \Delta_g\big(u_{i-1}(c(s), y)\big) s^{i-1} ds.$$

Remark. Here, we evaluate the geodesic c in terms of distance instead of "curve parameter" t as usual. We can do this because of the fact that geodesics are unique and minimizing within injectivity radii (i.e. outside of the cutlocus). For example, say that $x = c(a)$ and $y = c(b)$, then we have $c(0) := c(a), c\big(d(x, y)\big) := c(b)$ and $c\big(d(x, y)/2\big) := c\big((a + b)/2\big)$.

Lemma 4.24. The u_i are smooth on V_ϵ for all $0 \le i \le k$.

Proof. First, we have that u_0 is smooth, since by Lemma 4.23 we have that all the F_x are smooth on $B_\epsilon(x)$. Therefore, $\Delta_g u_0$ is smooth as well. Then, for u_1, plug in u_0 in the integral (which is smooth on V_ϵ), so that u_1 is smooth. Note that u_1 does not have a singularity in $x = y$. If we let $y \to x$, we have that $F_x(y) \to 1$. Since $\Delta_g u_0$ is smooth, it must have no singularities on V_ϵ. Thus, the resulting integral is proportional a power of $d(x, y)$ of at least 1 and the factor $(d(x, y))^{-1}$ is cancelled out. A similar argument goes for u_2 using the obtained smoothness of u_1. By repeating inductively, we see that all the u_i are smooth. □

Let us summarize what we have shown so far:

Proposition 4.25. The set of functions $u_k \in C^\infty(V_\epsilon), k \in \mathbb{N}$ defined by the recursion

$$u_0(x, y) = F_x^{-1/2}(y)$$
$$u_k(x, y) = (d(x, y))^{-k} F_x^{-1/2}(y) \int_0^r F_x^{1/2}\big(c(s)\big) \cdot \Delta_g\big(u_{k-1}(c(s), y)\big) s^{k-1} ds.$$

satisfies

$$L_y S_k = G \cdot t^k \cdot \Delta_g u_k$$

for all $k \in \mathbb{N}$, where $L_y = \Delta_g - \partial_t$ (where Δ_g acts on variable y). Moreover, we have that $u_0(x, x) = 1$ and by plugging this into the recursive formula we get $u_1(x, x) = R_g(x)/6$ where $R_g(x)$ defines the scalar curvature of (M, g) at x (how much is the space bent? This is a good question when studying, for instance, gravitation in spacetime. We will not go into this here).

Thus, we have established a function on all balls in M with the largest radius so that the Riemannian exponential function is a diffeomorphism on the balls. This function solves the heat equation in the limit of short time t. The next item on our schedule towards the fundamental solution is to extend the definition of S_k to all of M instead

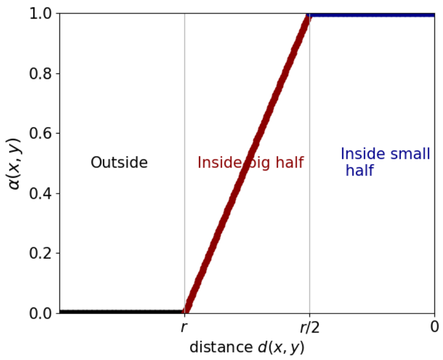

Figure 4.1: Bump function. The bump function $\alpha(x, y)$ is zero outside of the ϵ-ball and one inside of the $\epsilon/2$-ball around x (the small half of the ball). In between (inside of the big half of the ball), α goes from 0 to 1 in a smooth (in this case linear) manner.

of V_ϵ. To do this, we can employ a "partition of unity"-like function – a "bump" function that is zero everywhere except on V_ϵ. Then, at the edge, we let the function continuously increase in value until it reaches $V_{\epsilon/2}$, where it will be 1 uniformly. In other words, we define the function $\alpha \in C^\infty(M \times M, [0, 1])$ as

$$\alpha(x, y) := \begin{cases} 1 & y \in B_{\epsilon/2}(x) \\ D(x, y) & y \in B_\epsilon(x) \backslash B_{\epsilon/2}(x) \\ 0 & y \in M \backslash B_\epsilon(x) \end{cases} \tag{4.13}$$

where $D(x, y)$ is any smooth, strictly monotone decreasing function (with respect to $d(x, y)$) so that $D(x, y) \to 1$ if $d(x, y) \to \epsilon/2$ and $D(x, y) \to 0$ if $d(x, y) \to \epsilon$. An example of such a bump function is shown in Figure 4.1.
We get that $\alpha \equiv 0$ on $M \backslash V_\epsilon$ and $\alpha \equiv 1$ on $V_{\epsilon/2}$. Then, we define the function $H_k : M \times M \times (0, \infty) \to \mathbb{R}$ by

$$H_k(x, y, t) := \alpha(x, y) S_k(x, y, t).$$

This function is actually an extension of S_k from V_ϵ to M (technically, from $V_{\epsilon/2}$ to M) and has some interesting properties hinting towards the fundamental solution.

Lemma 4.26. Let $k > n/2$. Then, the function $H_k \in C^\infty(M \times M \times (0, \infty))$ has the following properties:

(i) $L_y H_k \in C^l(M \times M \times [0, \infty))$ for all $0 \leq l < k - n/2$,

(ii) For all $x, y \in M$, $H_k(x, y, t) \to \delta_x(y)$ as $t \to 0$.

(iii) Let $\tau > 0$. There exists a constant $A = A_\tau$ such that

$$\|(L_y H_k)(\cdot, \cdot, t)\|_{L^\infty(M \times M)} \leq A t^{k-n/2}$$

for all $t \in [0, \tau]$.

Proof. We are going to assume $l = 0$, the other cases follow (since the extension of $L_y H_k$ to $t = 0$, as we will see, vanishes).

(i) Now, the issue here is that we have to show continuity of $L_y H_k$ in $t = 0$. This is easy for $V_{\epsilon/2}$, because $\alpha \equiv 1$ there. Thus, $L_y H_k = G \cdot t^k \cdot \Delta_g(u_k) \to 0$ as $t \to 0$. Thus, continuity of $L_y H_k$ is preserved on $V_{\epsilon/2}$ if we set $L_y H_k = 0$ there for $t = 0$, i.e. $L_y H_k \equiv 0$ on $V_{\epsilon/2} \times \{0\}$. For $M \backslash V_\epsilon$ it is even easier, since $\alpha \equiv 0$ here and thus $H_k \equiv 0$ so we set $L_y H_k = 0$ on $t = 0$ and continuity is preserved in a similar way.
Finally we extend $L_y H_k$ to $t = 0$ on the parts in between, $V_\epsilon \backslash V_{\epsilon/2}$. Here, we have by Equation (4.5)

$$L_y H_k = \alpha \cdot (\Delta_g - \partial_t) S_k + \Delta_g(\alpha) \cdot S_k + 2g\big(\text{grad } \alpha, \text{grad } S_k\big)$$

which can be written as $G(x, y, t) \cdot f(x, y, t)$ for some function $f \in C^\infty(M \times M \times (0, \infty))$. This function has a singularity in $t = 0$ coming from grad S_k. That is, we claim that grad $S_k(x, y, t) \propto 1/t$ (i.e. has a pole of order 1). Indeed,

$$\text{grad } S_k = G \cdot \text{grad}\left(\sum_{i=0}^{k} u_i t^i\right) + (\text{grad } G) \cdot \sum_{i=0}^{k} u_i t^i$$

and we see that the first term yields only nonnegative powers of t, while grad $G \propto G/t$ as a result of the exponent. Since $d(x, y) > \epsilon/2 > 0$ for all $\epsilon > 0$ on this domain, we have that $G(x, y, t) \cdot f(x, y, t) \to 0$ as $t \to 0$ (due to the exponential factor) and thus we may also set $H_k \equiv 0$ at $t = 0$ here.

(ii) Let $f \in C^\infty(M)$ so that it is bounded and smooth. Remember that $H_k = \alpha G(u_0 + \ldots + t^k u_k)$. Hence, we want to prove that

$$\begin{aligned} f(x) &= \lim_{t\to 0} \sum_{j=0}^{k} t^j \int_M \alpha(x, y) G(x, y, t) u_j(x, y) f(y) dV_g(y) \\ &= \lim_{t\to 0} \sum_{j=0}^{k} t^j \Bigg(\int_{B_{\epsilon/2}(x)} \alpha(x, y) G(x, y, t) u_j(x, y) f(y) dV_g(y) \\ &\qquad + \int_{M \backslash B_\epsilon(x)} \alpha(x, y) G(x, y, t) u_j(x, y) f(y) dV_g(y) \Bigg). \end{aligned}$$

We look at this problem for each $0 \le j \le k$ separately. Since, for all $\epsilon > 0$ we have that $d(x, y) > \epsilon/2$ on the integration domain of the second term of the last expression, we see that the exponential factor in G makes the integrand, and therefore the entire term, vanish as $t \to 0$ (since the u_j are smooth). For the first term, we have on its integration domain $\alpha \equiv 1$. We write out $G(x, y, t)$ and we transform to normal coordinates instead of geodesic polar coordinates, i.e. $y \to v$ with $y = \exp_x(v)$, so that we integrate over the tangent space at x, $T_x M$, which, as we know, is isomorphic to $\mathbb{R}^n$. Denoting $J(v)$ for the associated Jacobian, we get that this term equals (omitting the summation over j and the t^j):

$$\begin{aligned} &\int_{B_{\epsilon/2}(x)} \frac{1}{(4\pi t)^{n/2}} e^{-d(x,y)^2/4t} u_j(x, y) f(y) dV_g(y) \\ &\qquad = \int_{T_x M} \frac{1}{(4\pi t)^{n/2}} e^{-\|v\|^2/4t} u_j(x, \exp_x(v)) f((\exp_x(v)) J(v) dv. \end{aligned}$$

Now comes a handy insight: On $T_xM \cong \mathbb{R}^n$, $\frac{1}{(4\pi t)^{n/2}}e^{-\|v\|^2/4t}$ is precisely equal to the fundamental solution $p(0,v,t)$ which approaches $\delta_0(v)$ as $t \to 0$. Therefore, the above expression approaches $u_j(x, \exp_x(0))f(\exp_x(0))J(0)$ and by seeing that $(\exp_x(0) = x)$ and $J(0) = 1$ (since there is no change in structure here at 0), we obtain that

$$\int_M \alpha(x,y)G(x,y,t)u_j(x,y)f(y)dV_g(y) \to u_j(x,x)f(x) \quad (t \to 0).$$

Now, using that $u_0(x,x) = 1$ we get

$$\int_M \alpha(x,y)G(x,y,t)u_0(x,y)f(y)dV_g(y) \to f(x) \quad (t \to 0)$$

and for any other $1 \leq j \leq k$ we have

$$t^j \int_M \alpha(x,y)G(x,y,t)u_j(x,y)f(y)dV_g(y) \to 0 \quad (t \to 0)$$

and hence

$$\sum_{j=0}^{k} \int_M \alpha(x,y)G(x,y,t)u_j(x,y)f(y)dV_g(y) = \int_M H(x,y,t)f(y)dV_g(y) \to f(x) \ (t \to 0).$$

Thus we may conclude that $H_k(x,y,t) \to \delta_x(y)$ as $t \to 0$.

(iii) First, note that the assertion holds for $t = 0$ by item (i). On $V_{\epsilon/2}$ it is clear, since here $L_yH_k = G \cdot t^k \cdot \Delta_g(u_k)$ and hence we can choose, for fixed t, $A = \sup_{(x,y)\in(M\times M)} G(x,y,t)\Delta_g(u_k(x,y,t)$ (which we can do since G and $\Delta_g u_k$ are smooth). Remember that G has a factor $t^{-n/2}$. On $M\backslash V_\epsilon$ we have $L_yH_k \equiv 0$ so the assertion holds. From the proof of item (i) we see that, on $V_\epsilon \backslash V_{\epsilon/2}$, $L_yH_k(x,y,t) = G(x,y,t)\cdot f(x,y,t)$ with for a smooth f. Here, we have $L_yH_k \to 0$ $(t \to 0)$ as well. Choose a $\tau > 0$. Since $S_k \to 0$ like $t^{k-n/2}$ as $t \to 0$, there must exist a constant so that this is true for all $t \in [0,\tau]$ (given that $L_yH_k \in C^l(M \times M \times [0,\infty))$.

□

So far, we have modified G to the extent that we have an expression H_k that approaches the Dirac delta for small t and that is defined on all of $M \times M \times [0,\infty)$. In addition, it solves the heat equation but only for $t = 0$ (by the proof of Lemma 4.26(i)). Thus, this is not yet a fundamental solution. At this point, we are going to need some understanding of how the heat operator L_y acts on the convolution of functions. Remember that for F, H continuous and integrable functions on $M \times M \times (0,\infty)$ we have the convolution

$$(F * H)(x,y,t) := \int_0^t \int_M F(x,z,s)H(z,y,t-s)dV_g(z)ds$$

from which it follows that for such an F, we have that $F * H_k \in C^l(M \times M \times (0,\infty))$ for all $0 \leq l < k - n/2$. Of course, this only makes sense if H_k is integrable. Showing this is very subtle and we will not do this here, but it follows from the fact that the u_j and G are integrable.
Let us find out how the heat operator L_y acts on $(F * H_k)(x,y,t)$.

Lemma 4.27. Let $F \in C^0(M \times M \times [0,\infty))$. Then

$$L_y(F * H_k)(x,y,t) = F(x,y,t) + \big(F * (L_y H_k)\big)(x,y,t).$$

Proof. Let us first determine the temporal partial derivative. We have

$$\begin{aligned}\partial_t(F * H_k)(x,y,t) &= \partial_t \int_0^t \int_M F(x,z,s) H_k(z,y,t-s) dV_g(z) ds \\ &\overset{\text{(a)}}{=} \lim_{s\to t} \int_M F(x,z,s) H_k(z,y,t-s) dV_g(z) + \int_0^t \int_M F(x,z,s)\partial_t\big(H_k(z,y,t-s)\big) dV_g(z) ds \\ &\overset{\text{(b)}}{=} F(x,y,t) + \int_0^t \int_M F(x,z,s)\partial_t\big(H_k(z,y,t-s)\big) dV_g(z) ds.\end{aligned}$$

Here, we used the Leibniz rule for integration at (a). At (b), we used the fact that s approaching t means $t-s$ approaching zero and then applying Lemma 4.26(ii) so that $H_k(z,y,t-s) \to \delta_z(y)$ and hence the integration cancels and we replace z with y in $F(x,z,t)$.
Now we can compute

$$\begin{aligned}L_y(F * H_k)(x,y,t) &= F(x,y,t) + \int_0^t \int_M F(x,z,s)\partial_t\big(H_k(z,y,t-s)\big) dV_g(z) ds \\ &\quad - \Delta_g \int_0^t \int_M F(x,z,s) H_k(z,y,t-s) dV_g(z) ds. \\ &\overset{*}{=} F(x,y,t) + \int_0^t \int_M F(x,z,s)(\partial_t - \Delta_g)\big(H_k(z,y,t-s)\big) dV_g(z) ds \\ &= F(x,y,t) + \int_0^t \int_M F(x,z,s) L_y\big(H_k(z,y,t-s)\big) dV_g(z) ds \\ &= F(x,y,t) + \big(F * (L_y H_k)\big)(x,y,t).\end{aligned}$$

where we used at $*$ that the integral is over z, not y, and that $F(x,z,s)$ is independent of y, so Δ_g acts only on $H_k(z,y,t)$ and thus we can write it inside of the integral and hence combine the two integrals into one. □

At last, we are ready to find a fundamental solution! We search for fundamental solutions of the form

$$p = H_k - F * H_k$$

for a suitable choice of F. Let us introduce a final piece of notation.

Notation 4.28. For continuous, square-integrable maps F and $j = 1, 2, \dots$ we write

$$(F)^{*j} := F \underbrace{* \dots *}_{j \text{ times}} F.$$

If $j = 1$ we have $(F)^{*1} = F$.

Since we want $L_y p$ to vanish, we can consider for all $k > n/2$ the functions

$$F_k := \sum_{j=1}^{\infty} (-1)^{j+1} (L_y H_k)^{*j}.$$

Let us assume for now that this series converges (as we will show shortly). Then $p = H_k - F_k * H_k$ solves the heat equation, as we will see now.

Lemma 4.29. Let F_k be as defined above. Then $p := H_k - F_k * H_k$ is a solution to the heat equation.

Proof. First notice that, for such a p,

$$\begin{aligned} L_y p &= L_y(H_k - F_k * H_k) \\ &= L_y H_k - L_y(F_k * H_k) \\ &= L_y H_k - F_k - \big(F_k * (L_y H_k)\big). \end{aligned} \tag{4.14}$$

Then, by Equation (4.14) we find that

$$\begin{aligned} L_y p &= L_y H_k - F_k - \big(F_k * (L_y H_k)\big) \\ &= L_y H_k - \sum_{j=1}^{\infty}(-1)^{j+1}(L_y H_k)^{*j} - \left(\sum_{j=1}^{\infty}(-1)^{j+1}(L_y H_k)^{*j}\right) * (L_y H_k) \\ &= L_y H_k - \sum_{j=1}^{\infty}(-1)^{j+1}(L_y H_k)^{*j} - \left(\sum_{j=1}^{\infty}(-1)^{j+1}(L_y H_k)^{*j+1}\right) \\ &= L_y H_k - \sum_{j=1}^{\infty}(-1)^{j+1}(L_y H_k)^{*j} - \left(\sum_{j=1}^{\infty}(-1)^{j+1}(L_y H_k)^{*j} + L_y H_k\right) = 0. \end{aligned} \tag{4.15}$$

□

Proposition 4.30. Let $0 \le l < k - n/2$. Then the series $F_k \in C^l(M \times M \times [0,\infty))$. In particular, it is convergent. Moreover, let $\tau > 0$. Then there exists a constant $C = C_\tau$ such that

$$\|F_k(\cdot,\cdot,t)\|_{L^\infty(M\times M)} \le C t^{k-n/2} \qquad \text{for all } t \in [0,\tau]$$

where $\|F_k(\cdot,\cdot,t)\|_{L^\infty(M\times M)} := \sup_{(x,y)\in(M\times M)}(F_k(x,y,t))$.

Proof. By Lemma 4.26(iii), we have that there must exist a constant $A = A_\tau$ such that

$$\|(L_y H_k)(\cdot,\cdot,t)\|_{L^\infty(M\times M)} \le A t^{k-n/2}$$

for all $t \in [0,\tau]$, in particular for $t = \tau$. Let $\int_M dV_g(y) =: \mathrm{vol}_g(M)$. Then, for all $j = 1, 2, \ldots$ and $t \in [0,\tau]$ we have [6]:

$$\|(L_y H_k)^{*j}(\cdot,\cdot,t)\|_{L^\infty(M\times M)} \le \frac{A(A\tau^{k-n/2})^{j-1}\mathrm{vol}_g(M)^{j-1}}{(k-\frac{n}{2}+j-1)\cdots(k-\frac{n}{2}+2)(k-\frac{n}{2}+1)} t^{k-\frac{n}{2}+j-1}.$$

We can apply the ratio test to this to find that

$$\lim_{j\to\infty} \frac{\|(L_y H_k)^{*j+1}(\cdot,\cdot,t)\|_{L^\infty(M\times M)}}{\|(L_y H_k)^{*j}(\cdot,\cdot,t)\|_{L^\infty(M\times M)}} \le \lim_{j\to\infty} \frac{A\tau^{k-n/2}\mathrm{vol}_g(M)}{(k-\frac{n}{2}+j)} t = 0 < 1$$

for all $k > n/2$. Thus, we find that $\sum_{j=1}^{\infty}(L_y H_k)^{*j}$ converges absolutely uniformly (w.r.t. the L^∞-norm) if $k > n/2$, which gives us the second assertion. Moreover, it gives that $\sum_{j=1}^{\infty}(-1)^{j+1})(L_y H_k)^{*j} = F_k$ converges uniformly, i.e. to a continuous function. By functional analysis statements we conclude that it must be C^∞ but since $L_y H_k$ is only C^l, so is F_k. □

Now, since we have proved that F_k is well-defined, it is time that we prove the main result of this chapter.

Theorem 4.31. *For compact Riemannian manifold (M, g) the fundamental solution of the heat equation exists and is given by*

$$p = H_k - F_k * H_k$$

for all $k > n/2 + 2$.

Proof. By Lemma 4.29 we know that p is a solution to the heat equation. Moreover, since $k > n/2 + 2$, we have that $p \in C^2(M \times M \times (0, \infty))$. Next, we prove that $p(x, y, t) \to \delta_x(y)$ as $t \to 0$. By Lemma 4.26 we have that this is the case for H_k, thus:

$$\begin{aligned} \lim_{t\to 0} \int_M p(x, y, t) f(y) dV_g(y) &= \lim_{t\to 0} \int_M \big(H_k(x, y, t) - (F_k * H_k)(x, y, t)\big) f(y) dV_g(y) \\ &= f(x) - \lim_{t\to 0} \int_M (F_k * H_k)(x, y, t) f(y) dV_g(y). \end{aligned} \tag{4.16}$$

Now, let C so that by Proposition 4.30 we have $\|F_k(\cdot, \cdot, t)\|_{L^\infty(M\times M)} \leq Ct^{k-n/2}$ for all $t \in [0, \tau]$. Then, the kernel $R_k := t^{-(k-n/2)} F_k$ is (uniformly) bounded on $M \times M \times [0, \tau]$ (its L^∞-norm must be at most C). Thus, we have that, for $t \in [0, \tau]$, $R_k * H_k(x, y, t) f(y)$ is bounded as well, hence so is its integral over M. Therefore, we obtain (for $k > n/2$ thus in particular $k > n/2 + 2$):

$$\lim_{t\to 0} \int_M (F_k * H_k)(x, y, t) f(y) dV_g(y) = \lim_{t\to 0} t^{k-n/2} \int_M (R_k * H_k)(x, y, t) f(y) dV_g(y). = 0.$$

Thus, by Equation (4.16) we have that

$$\lim_{t\to 0} \int_M p(x, y, t) f(y) dV_g(y) = f(x)$$

and therefore, $p(x, y, t) \to \delta_x(y)$ as $t \to 0$. □

This concludes the main result of this chapter: for any compact Riemannian manifold there must exist a fundamental solution to the heat equation – a solution with a point source as initial condition. In addition, by Proposition 4.13, it must be unique.

Collorary 4.32. The fundamental solution $p = H_k - F_k * H_k$ is unique, and hence it is (or, can be defined) independent of k.

Let us look at an example on a simple space. First, we give an important result.

Proposition 4.33. (without proof) [6]. Let (N, g_N) be a Riemannian manifold and let Γ be a discrete algebraic group acting properly and freely on N. Let (M, g_M) be the resulting compact quotient manifold $M = N/\Gamma$ (Since M is compact, the group Γ is finitely generated, thus Γ is a countable group). Let $p_N \in C^\infty(N \times N \times (0, \infty))$ be the fundamental solution on N given by $p_N = H_k - F_k * H_k$ as before. Then the fundamental solution for the heat equation on (M, g_M) is given by $p : M \times M \times (0, \infty) \to \mathbb{R}$,

$$p(x, y, t) := \sum_{\gamma\in\Gamma} p_N(x_N, \gamma \cdot y_N, t)$$

where $\cdot$ denotes the group operation of Γ.

Example 4.34. Let the fundamental solution on $\mathbb{R}^n$ be given by Equation (4.10). We look at the Torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$. Here, the group operation of $\mathbb{Z}^n$ is component-wise addition. Let the corresponding covering map be

$$\pi : (\mathbb{R}^n, g_{\text{Eucl}}) \to (\mathbb{T}^n, g_{\mathbb{T}})$$

and define $x', y' \in \mathbb{R}^n$ so that $\pi(x') = x \in \mathbb{T}$ and $\pi(y') = y \in \mathbb{T}$. Then we have that the fundamental solution on $\mathbb{T}^n$ is given by

$$p_{\mathbb{T}}(x, y, t) := \frac{1}{(4\pi t)^{n/2}} \sum_{m \in \mathbb{Z}^n} e^{-\|y'+m-x'\|^2/4t}.$$

For the one-dimensional torus, that is, the circle S^1, we find

$$p_{S^1}(x, y, t) = 2\sqrt{\pi t}\theta_3\bigg((x-y)\pi, e^{-4\pi^2 t}\bigg)$$

where $\theta_3(p, q)$, where $\theta_3(p, q)$ is the "third order" Jacobi theta function (Riemann himself came up with this notation). This example illustrates that the fundamental solution is hard to explicitly define, even for simple spaces.

This concludes the chapter on the heat equation. Next, we discuss an application in Riemannian geometry and stochastic analysis – Brownian motion.

5. Brownian Motion

In this chapter, we discuss how Brownian motion behaves on Riemannian manifolds. To this end, we first introduce the concept of Brownian motion in general. Before jumping to Riemannian manifolds, the case of $\mathbb{R}^n$ is discussed. Similar to Chapter 4, we have in mind a main statement that we want to prove. This statement is that the *transition density function* of the Brownian motion is in fact equal to the fundamental solution of the heat equation (albeit the heat equation with an extra factor of one half). The transition density function can be interpreted as follows: if we select a region $D \subset M$ and a starting point $p \in M$ at $t = 0$. What is the probability that after time t, the Brownian motion will be inside of the region D? It certainly is convenient that the fundamental solution integrates to unity. But more on that later. We assume that the reader is familiar with basic probability theory and stochastic processes.

5.1 Definitions

Brownian motion is a concept that is often seen in physics literature. It can be interpreted as the random movement of a particle suspended in some medium. For instance, Einstein studied this process in the context of the *diffusion* of gas particles. The diffusion of gas particles is related to the mean square distance that a particle travels in a certain time. Therefore, Brownian motion is often related to discrete spaces as a "random walk": at any discrete time interval Δt, the particle has a finite amount of options for a movement and each movement has a certain finite probability. For continuous spaces, each movement has probability zero, but we can define a probability density so that there is a finite chance that the particle will move to a certain region. In a one-dimensional ideal gas, Einstein figured out that this probability density with initial position $y \in \mathbb{R}$ is actually a Gaussian distribution with mean y and variance $2Dt$ with D the diffusion coefficient for the gas. This seems reasonable: on average, the particle travels further at larger time intervals.
Formally, we can define "a Brownian motion" as a stochastic process that encompasses the above discussion. Before defining this, we remember a few definitions from probability theory.

Definition 5.1. A **probability space** is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ that provides a formal model of random processes. It contains three elements:

(i) A *sample space* Ω which is a set that contains all possible outcomes of the process, such as "the value of the thrown die is 3."

(ii) An *event space* $\mathcal{F}$ which is a set of events. An event is a set of outcomes in the sample space, such as "the value of the thrown die is less than 3."

(iii) A *probability function* $\mathbb{P} : \mathcal{F} \to [0, 1]$ which assigns a certain probability between 0 and 1 to each event in the event space, such as "the probability that the die lands on 3 is 1/6" or "the probability that the value of the thrown die is less than three is 1/3."

Definition 5.2. A **random variable** is a variable whose values depend on outcomes of a random process. Formally, it is a measurable function $X : \Omega \to \mathbb{R}$ defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It represents the possible outcomes of an experiment or process.

Let us state the definition of a stochastic process as a brief reminder.

Definition 5.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then a **stochastic process** is a set of random variables $\{X_j\}_{j \in J}$ for some indexing set J. A simple example is repeatedly flipping a coin and noting whether it lands on heads or tails (this is a Bernoulli process).

The following discussion contains insights from [9].

Definition 5.4. Let $\{X_t\}_{t \geq 0}$ be a stochastic process indexed by the nonnegative real numbers t. We say that it has **independent increments** if for all choices of real numbers

$$0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \ldots \leq s_n < t_n < \infty$$

we have that the **increment variables**

$$X_{t_1} - X_{s_1}, \ldots, X_{t_n} - X_{s_n}$$

are all independent of each other. Additionally, it has **stationary increments** if for any $0 < s, t < \infty$ we have that the distribution of increment variable $X_{t+s} - X_s$ has the same distribution as $X_t - X_0$.

Definition 5.5. (Standard Brownian motion). A **Standard Brownian motion** is a stochastic process $\{X_t\}_{t \geq 0}$ indexed by the nonnegative real numbers t, where the X_t take values to $\mathbb{R}$ that satisfies the following properties:

(i) $X_0 = 0$,

(ii) The function $t \mapsto X_t$ is continuous,

(iii) The process $\{X_t\}_{t \geq 0}$ has stationary, independent increments,

(iv) The increment $X_{t+s} - X_s$ has a Gaussian distribution around 0 with variance t.

Formally, Brownian motion is defined on the probability space $(\mathbb{R}, \mathcal{P}(\mathbb{R}), \mathbb{P})$ where p is determined by the *transition density function* defined at the end of this section. In SECTION 5.2, this definition will be extended to $\mathbb{R}^n$, and to Riemannian manifolds in general in SECTION 5.3.

We can interpret this as follows:

(i) At $t = 0$ the particle cannot have travelled anywhere, hence the process X_0 represents a displacement of 0.

(ii) The "movement" of the particle is continuous, i.e. it cannot make any discontinuous jumps, hence it follows a continuous curve.

(iii) Independence means that once the particle has moved along a certain path at time t_1, this path has no influence whatsoever on the path it will follow after time t_1 (hence it can be interpreted as "fully random motion independent of the past"). Stationary increments say that for any $0 < s, t < \infty$ we have that the position at time $t + s$ has the same probability density in the motion starting from time s as that the position at time t has while starting from time 0 (since $X_t - X_0 = X_t$).

(iv) The probability density of motion starting from time s and ending in time $s+t$ is determined by the Gaussian probability density. By (iii), it is also true for $s = 0$.

Brownian motions are also called *Wiener processes* since Wiener proved that such a stochastic process always exists on a given probability space.
In the following, we state some definitions on probability theory. We will not go in full detail for some definitions, but they are necessary for proving the main result of this chapter.

Definition 5.6. Let X be a set and let $\mathcal{P}(X)$ be its power set. Then a σ**-algebra** is a set $\Sigma \in \mathcal{P}(X)$ so that

(i) Σ is closed under complementation (i.e. if $S \in \Sigma$, then $X \backslash S \in \Sigma$),

(ii) Σ contains X itself and the empty set,

(iii) Σ is closed under countable unions (and hence by (i), under countable intersections).

Definition 5.7. Let X be a set. Let $F \subset \mathcal{P}(X)$ be an arbitrary family of subsets of X. Let $\Sigma_F := \{\Sigma \mid \Sigma \text{ is a } \sigma\text{-algebra containing } F\}$. Then the σ**-algebra generated by** F is given by

$$\sigma(F) := \bigcap_{\Sigma \in \Sigma_F} \Sigma,$$

which is the unique smallest σ-algebra containing F.

Definition 5.8. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let J be a countable index set. For each $j \in J$, let $\mathcal{F}_j$ be a σ-algebra of $\mathcal{F}$. Then

$$\mathcal{F}_* := (\mathcal{F}_j)_{j \in J}$$

is called a **filtration** of $\mathcal{F}$ if $\mathcal{F}_i \subset F_j$ for all $i \leq j$. The space $(\Omega, \mathcal{F}, F_*, \mathbb{P})$ is then called a **filtered probability space**. For example, let $\{X_n\}_{n \in \mathbb{N}}$ be a stochastic process on the probability space. Then let $\mathcal{F}_n := \sigma(X_k | k \leq n)$ (that is, the σ-algebra generated by the random variables $X_1, \ldots, X_n$). Then $(\mathcal{F}_n)_{n \in N}$ is a filtration.
Heuristically, we can interpret filtrations as a set of information. The amount of information then increases as the stochastic process progresses. In the example above, $\mathcal{F}_n$ would be the information in the first n random variables.

For Brownian motion we often work with probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a specific filtration $\mathcal{F}_* := (\mathcal{F}_t)_{t \geq 0}$ containing information on the motion during the time interval $[0, t]$.

Definition 5.9. The independence of Brownian motion from its past trajectory ("memoryless") is a result of the **Markov property** which says that, for bounded continuous maps f on a certain region D and for a Brownian motion $\{X_t\}_{t \geq 0}$ on a probability space with filtration $\mathcal{F}_* = (\mathcal{F}_t)_{t \geq 0}$, we must have that $\mathbb{E}(f(X_t) \,| \mathcal{F}_s) = \mathbb{E}(f(X_t) \,| \sigma(X_s))$ where $\mathbb{E}$ denotes the expectation value with respect to probability function $\mathbb{P}$.

Next, we introduce a notion of *explosion time* of a stochastic process. In many one-dimensional ordinary differential equations of the form

$$\partial_t u = b(u), \quad u(0) = x_0$$

for some function b so that $b(u) > 0$ for all u, there exists a finite time T such that the solution $u(t)$ blows up as $t \to T$, i.e. $\lim_{t \to T} u(t) = \infty$ [10]. Technically, this is the case if and only if $\int_0^\infty 1/b(u(t))dt < \infty$. In that case we can find an exact solution for T. As was shown in [10], although we cannot find an exact solution, there is some finite time at which solutions to *stochastic* differential equations (such as the heat equation on Brownian motion) "explode".

Definition 5.10. A **stochastic differential equation** is a differential equation in which one or more terms is a stochastic process. As a result, the solution is also a stochastic process. A typical stochastic equation is of the form

$$dY_t = \mu(Y_t, t)dt + \sigma(Y_t, t)dX_t$$

where μ and σ are some functions of the stochastic process $Y_t := \{Y_t\}_{t \geq 0}$ and where $X_t := \{X_t\}_{t \geq 0}$ denotes a standard Brownian motion. Heuristically, this stochastic differential equation tells us that the change in Y_t over an infinitesimally small time interval $[t, t + dt]$ is normally distributed with mean $\mu(Y_t, t)$ and variance $\sigma(Y_t, t)^2$.

Definition 5.11. The finite (positive) time at which the solution to a stochastic differential equation approaches infinity is the **explosion time** and is denoted by e.

Definition 5.12. Let $(\Omega, \mathcal{F}, \mathcal{F}_*, \mathbb{P})$ be a filtered probability space with filtration $\mathcal{F}_* = (\mathcal{F}_t)_{t \geq 0}$. Let τ be a random variable with a strictly positive value. Then τ is called a **stopping time** with respect to the filtration $\mathcal{F}_*$ if

$$\{\tau \leq t\} \in \mathcal{F}_t$$

for all $t \geq 0$. Essentially, what this requirement says is that along with the filtration comes an explicit stopping time. For example, if you flip a coin five times, then $\tau = 5$ provides a stopping time. In Brownian motion, let $D \subset \mathbb{R}$ be a specific region (later we replace $\mathbb{R}$ with $\mathbb{R}^n$ or a Riemannian manifold), then τ_D can be set the minimal time during which the Brownian motion stays within the region D. Hence, it is a stopping time as well. Formally, this means that we create a new random variable by mapping the first random variable X_t in the Brownian motion $\{X_t\}_{t \geq 0}$ for which it is not in D to the associated time t.
A non-example of a stopping time is flipping a coin until the total amount of heads is 20 more than the total amount of tails, since this does not provide an explicit time at which we can stop flipping the coin. To know a "stopping time" we require, at some point, knowledge about the past, present *and* future. Therefore, it does not satisfy the requirement on the filtration above.

Notation 5.13. (wedge notation). We write for $t, s \in \mathbb{R}$:

$$t \wedge s := \min(s, t).$$

For example, if we want to study Brownian motion in a certain region D, we can look at probabilities for times $t \wedge \tau_D$, that is, for all times up until and including the moment that the Brownian motion leaves the region D.

Next, we introduce the notion of a *martingale*. In essence, it is a sequence of stochastic processes for which, at a given point in time, the conditional probability of the next value in the sequence is precisely equal to the present one (the condition being "given all other values for smaller times", and these values have no influence). Therefore, we can interpret a martingale as a stochastic process for a "fair game". Let us formally define martingales in the context of Brownian motions.

Definition 5.14. Let $\{X_t\}_{t\geq 0}$ be a Brownian motion. Then a **continuous-time martingale** with respect to the filtration $\mathcal{F}_* = (\mathcal{F}_t)_{t\geq 0}$ of $\{X_t\}_{t\geq 0}$ is a stochastic process $\{Y_t\}_{t\geq 0}$ such that for all t we have

$$\mathbb{E}(|Y_t|) < \infty$$

and

$$\mathbb{E}(Y_t \mid \{X_T, T \leq s\}) = Y_s$$

for all $0 \leq s \leq t$. An example of a martingale is a Brownian motion, or a Brownian motion up to a certain stopping time ("stopped process"), i.e. $X_{t\wedge T}$.

We conclude this section with a formal definition of the probability that a Brownian motion will still be in a certain region D after time t, or in particular, the probability density in each given point (so that a probability will be given by integration over the probability density). However, in order to talk about Brownian motion on manifolds, we must first define this.

Definition 5.15. A **Brownian motion starting at** $x \in M$ for a Riemannian manifold (M, g) is a stochastic process $\{X_t^x\}_{t\geq 0}$ indexed by the nonnegative real numbers t, where the X_t^x take values to M. It satisfies properties (ii) and (iii) from Definition 5.5 and the following:

(i) $X_0^x = x$,

(ii) The increment $X_{t+s}^x - X_s^x$ has a probability distribution equal to the transition density function defined below.

(iii) Itô's Lemma is satisfied (we will discuss this in SECTION 5.3).

Definition 5.16. Let (M, g) be a Riemannian manifold. Let $X_t := \{X_t\}_{t\geq 0}$ be a Brownian motion with explosion time e. Let $D \subset M$ be a connected region of M. Let $x \in D$. Then we denote the probability measure on M for a Brownian motion starting at x by $\mathbb{P}_x$. Then, the **transition density function** $p_M(x, y, t)$ for Brownian motion is defined so that

$$\mathbb{P}_x(X_t \in D, t < e) = \int_D p_M(x, y, t)\, dV_g(y). \tag{5.1}$$

Recall that $dV_g(y)$ represents the Riemannian volume measure induced by the metric g in the point y.

5.2 Transition density in $\mathbb{R}^n$

Let $p_{\mathbb{R}^n}$ be the fundamental solution to the heat equation on $\mathbb{R}^n$ with a slightly different heat operator

$$L := \frac{1}{2}\Delta - \partial_t. \tag{5.2}$$

Then $p_{\mathbb{R}^n}$ is simply given by Equation (4.10) but with the substitution $t \to t/2$, that is,

$$p_{\mathbb{R}^n}(x, y, t) = \frac{1}{(2\pi t)^{n/2}} e^{-\|y-x\|^2/2t}. \tag{5.3}$$

Proposition 5.17. The transition density function $p_{\mathbb{R}^n}$ of Brownian motion on $\mathbb{R}^n$ as in Equation (5.1) is given by Equation (5.3).

Proof. For $n = 1$ this is obvious. Namely, the Brownian motion starting at x is defined to have a Gaussian distribution around x, which is given by Equation (5.3). We can extend this argument to $n > 1$. Since each dimension adds a "degree of freedom" to the Brownian motion's "particle", we see that it can move independently in all orthogonal directions (i.e. along the directions of the standard basis of $\mathbb{R}^n$). Hence, the displacement in one direction is independent from the displacement in the other. Therefore we must have that

$$\begin{aligned} p_{\mathbb{R}^n}(x, y, t) &= \prod_{i=1}^{n} p_{\mathbb{R}}(x_i, y_i, t) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi t}} e^{-(y_i - x_i)^2/2t} \\ &= \frac{1}{(2\pi t)^{n/2}} e^{-\left[\sum_{i=1}^{n}(y_i - x_i)^2\right]/2t} \\ &= \frac{1}{(2\pi t)^{n/2}} e^{-\|y-x\|^2/2t}. \end{aligned}$$

□

Remark. By Definition 5.15 we see that we can *define* Brownian motion on $\mathbb{R}^n$ to be a Brownian motion whose variables take values in $\mathbb{R}^n$ instead of $\mathbb{R}$ and whose increments $X_{t+s} - X_s$ have an n-dimensional Gaussian distribution around 0 with variance t.

5.3 Transition density on Riemannian manifolds

We will now prove that the fundamental solution of the heat equation is the transition density function of Brownian motion on all compact Riemannian manifolds (M, g). We follow mainly the discussion of [11]. First, we state the following important result (which we will not prove).

Lemma 5.18. (Itô's Lemma).[12] Let $f \in C^2(M)$. Let $X_t := \{X_t\}_{t \geq 0}$ be a Brownian motion on M. Then we have that

$$f(X_t) = f(X_0) + \text{ martingale } + \frac{1}{2}\int_0^t \big(\Delta_g(f)\big)(X_s)ds.$$

The martingale can be expressed explicitly on $\mathbb{R}^n$ but not in general on Riemannian manifolds.

While we prove the main result, we will generalize the existence and uniqueness of the fundamental solution on compact Riemannian manifolds with a boundary. During this section, we let (M, g) denote a compact Riemannian manifold (with boundary). Here, we denote the heat operator by

$$L_g := \frac{1}{2}\Delta_g - \partial_t. \tag{5.4}$$

Proposition 5.19. Let D be a smooth, relatively compact (i.e. its closure $\overline{D}$ is compact in M) domain on M. Then, there exists a unique fundamental solution of the heat equation with heat operator (5.4), $p_D(x,y,t)$, on $\overline{D} \times \overline{D} \times (0,\infty)$ such that

$$\int_D p_D(x,y,t)dV_g(y) \leq 1$$

where equality holds if $\overline{D} = M$ (a detailed discussion on this is given in SECTION 6.2). Moreover, it satisfies the Dirichlet boundary condition for each $x \in D$ fixed:

$$p_D(x,y,t) = 0, \quad x \in \partial D.$$

We call $p_D(x,y,t)$ the *Dirichlet heat kernel*.

Proof. This is a consequence of the proof in SECTION 4.5 since $\overline{D}$ is compact. For the latter assertion, the results in that section can be shown for compact manifolds with boundary using an adaptation of Duhamel's principle and the fact that for all $\varphi, \psi \in C^\infty(M)$ we must have that

$$\int_{\partial M} \varphi(y)\frac{\partial}{\partial x^j}\psi(y)dS_g(y) = 0$$

for all $1 \leq j \leq n$, where x is a chart and $dS_g(y)$ denotes a surface area element induced by the metric g, similar to how dV_g is a volume element induced by g. For an elaborate discussion, see Chavel [5]. □

When x or y is not in D we put $p_D(x,y,t) = 0$ to preserve continuity. We prove a consequence of the above.

Proposition 5.20. Let the setting be the same as in Proposition 5.19. Let f be a bounded, continuous map on D. Then the heat equation with initial and boundary conditions

$$\begin{cases} L_g u(x,t) = 0 & (x,t) \in \overline{D} \times (0,\infty) \\ u(x,0) = f(x) & x \in D \\ u(x,t) = 0 & (x,t) \in \partial D \times (0,\infty) \end{cases} \tag{5.5}$$

has the unique solution

$$u(x,t) = \int_D p_D(x,y,t)f(y)dy.$$

Proof. The fact that u is a solution to (5.5) follows immediately from Proposition 5.19. The proof for the uniqueness is precisely analogous to the proof of the uniqueness of the regular homogeneous heat equation (Lemma 4.10). □

Next, we are going to identify the Dirichlet heat kernel $p_D(x,y,t)$ as the transition density function of Brownian motion. That is, the Brownian motion that is "killed" at the moment the motion leaves the region D. This moment is given by the stopping time

$$\tau_D := \inf\{t > 0 | X_t \notin D\}.$$

We will refer to τ_D as the "first exit time" rather than stopping time, as it is in essence the moment in time when the Brownian motion exits D.

Proposition 5.21. Let $p_D(x,y,t)$ be the Dirichlet heat kernel for relatively compact domain $D \subset M$. Let the first exit time of Brownian motion on M from D be given by τ_D. Then we have that

$$\mathbb{P}_x(X_t \in D, t < \tau_D) = \int_D p_D(x,y,t)dV_g(y),$$

that is, p_D is the transition density function of Brownian motion on M that is killed at ∂D.

For the proof we are going to need the following definition.

Definition 5.22. An **exhaustion** of a topological space M is a sequence of relatively compact sets (i.e. their closure is compact in M) $\{D_n\}_{n\in\mathbb{N}}$ such that the closure of D_n is contained in the interior of D_{n+1} and we have $X = \bigcup_{n=1}^{\infty} D_n$, i.e. $\lim_{n\to\infty} D_n = M$.

Proof. (of Proposition 5.21). By Proposition 5.20 we have that the solution of the boundary condition version of the heat equation (5.5) is given by

$$u(x,t) = \int_D p_D(x,y,t)f(y)dV_g(y).$$

Next, let $0 < s < t \wedge \tau_D$. Then, we apply Itô's Lemma 5.18 (see the remark below this proof) to $u(X_s, t-s)$ to find

$$\begin{aligned} u(X_s, t-s) &= u(x,t) + \text{ martingale } + \int_0^s (\frac{1}{2}\Delta_g - \partial_t)u(X_s, t-r)dr \\ &= u(x,t) + \text{ martingale } + \int_0^s L_g u(X_s, t-r)dr. \end{aligned}$$

Here, the last term vanishes because u solves the heat equation. Now, let $\{t_n\}_{n\in\mathbb{N}}$ be a sequence such that $\lim_{n\to\infty} t_n = t$ and let $\{D_n\}_{n\in\mathbb{N}}$ be an exhaustion of D. Then, let $s = \lim_{n\to\infty} s_n = t_n \wedge \tau_{D_n}$ (we write $s = t \wedge \tau_D$ for simplicity). In addition, we take the expected value with respect to $\mathbb{P}_x$ (denoted $\mathbb{E}_x$). We obtain

$$\begin{aligned} \mathbb{E}_x\Big(u\big(X_{t\wedge\tau_D}, t-(t\wedge\tau_D)\big)\Big) &= \int_D u(y,t)p_D(x,y,t)dV_g(y) + 0 \\ &= u(x,t) \end{aligned}$$

where the martingale has vanished. Martingales have the property that if X_t is a martingale, then for all $t > 0$ we have that $\mathbb{E}(X_t) = \mathbb{E}(X_0)$. In this case, the martingale is a integral from 0 to t, and it must be equal to the case $t = 0$ which of course integrates to 0.
In [11] Hsu showed that the Markov property of Brownian motion leads to the fact that we can separate the expectation value of $f(X_t)$ into two terms: the expectation value of $u(x,t)$ up to the first exit time τ_D and the expectation value of $u(x,t)$ thereafter. Now, we have that u vanishes on (and thus beyond) the boundary of D. Thus, the second term vanishes and we have

$$\mathbb{E}_x(f(X_t), t < \tau_D) = \mathbb{E}_x\Big(u\big(X_{t\wedge\tau_D}, t-(t\wedge\tau_D)\big)\Big) = u(x,t) = \int_D p_D(x,y,t)f(y)dV_g(y).$$

From probability theory we see that this is equivalent to

$$\mathbb{P}_x(X_t \in D, t < \tau_D) = \int_D p_D(x,y,t)dV_g(y),$$

concluding the proof. □

Remark. In the above proof, we have applied a slightly more general version of Itô's Lemma, which also includes a time variable, accounting for the term $-\partial_t u(X_s, t-r)$ in the integral. For a detailed discussion, see the lecture notes on stochastic differential equations by Ruszel [13]. Thus, we almost have the main result. We are ready to prove it now.

Theorem 5.23. *Let p be the fundamental solution to the heat equation with heat operator L_g given by*

$$p = H_k - F * H_k$$

as in SECTION 4.5 (of course, the definition of $G(x, y, t)$ has slightly changed: similar to in $\mathbb{R}^n$, we have $t \to t/2$ and thus, so has $H_k := G \cdot S_k$). Then, p is the transition density function of Brownian motion on M.

Proof. Let $\{D_n\}_{n\in\mathbb{N}}$ be an exhaustion of M. Let $f \in C^0(M)$. Then, by the previous proof we have

$$\mathbb{E}_x(f(X_t), t < \tau_{D_n}) = \int_{D_n} p_{D_n}(x, y, t) f(y) dV_g(y) = \int_M p_{D_n}(x, y, t) f(y) dV_g(y) \quad (5.6)$$

since $p_{D_n}(x, y, t) = 0$ if x or y is not in D_n. Thus, we have

$$\int_M \Big[p_{D_{n+1}}(x, y, t) - p_{D_n}(x, y, t) \Big] f(y) dy$$
$$= \mathbb{E}_x(f(X_t), \tau_{D_n} \le t < \tau_{D_{n+1}}) \ge 0.$$

Hence, we have that $p_{D_{n+1}}(x, y, t) \ge p_{D_n}(x, y, t)$ on $M \times M \times (0, \tau_{D_{n+1}})$.
Now, let $n \to \infty$. We define

$$p_M(x, y, t) = \lim_{n\to\infty} p_{D_n}(x, y, t).$$

Since Brownian motion cannot leave M, there is no exit time τ_M. Therefore, the Brownian motion is limited only by explosion time e. Also, by Proposition 5.19 we have that $\int_M p_M(x, y, t) dV_g(y) = 1$. Since p_{D_n} is the fundamental solution on D_n we can conclude that p_M is a fundamental solution on M. By Equation (5.6) we have that, using the fact that the sequence $\{\tau_{D_n}\}_{n\in\mathbb{N}}$ defined in the previous proof now approaches e instead of some exit time as $n \to \infty$, we obtain

$$\mathbb{E}_x(f(X_t), t < e) = \int_M p_M(x, y, t) f(y) dV_g(y)$$

or, equivalently,

$$\mathbb{P}_x(X_t \in \tilde{D}, t < e) = \int_{\tilde{D}} p_M(x, y, t) dV_g(y)$$

for all $\tilde{D} \subset M$ (actually, on compact manifolds there is no escaping the manifold and thus $e = \infty$ with probability one, see SECTION 6.2). Thus p_M is the transition density function for Brownian motion on M.
Finally, by Lemma 4.13, the fundamental solution to the heat equation with L_g on M is unique and hence, we have that $p_M = p$. This shows the wanted result. □

Remark. By Definition 5.15 we see that we can *define* Brownian motion on M to be a Brownian motion whose variables take values in M instead of $\mathbb{R}$, satisfying Itô's Lemma and whose increments $X_{t+s} - X_s$ have a probability distribution given by the fundamental solution to the heat equation $L_g u = 0$.

This concludes the chapter on Brownian motion. However, the next chapter includes a summary of this thesis and will discuss some applications of this result and Brownian motion and the heat equation on Riemannian manifolds in general.

6. Applications & Outlook

In this chapter we look at a few geometric and probabilistic applications and/or lingering questions concerning the heat equation and Brownian motion on Riemannian manifolds. The first subject will be on the fundamental solutions and how it should be decaying as a function of geodesic distance (it turns out that this is very hard to prove for general cases). Then, we will discuss Brownian motion on non-compact Riemannian manifolds and how, in that case, Brownian motion can 'escape' from the manifold similar to how it could exit a region $D \subset M$ discussed in Chapter 5. This has to do with a property called *stochastic completeness*. Finally, we will discuss the behaviour of solutions to the heat equation and Brownian motion on a short time scale (called *short-time asymptotics*).

6.1 Fundamental solution and geodesic distance

6.1.1 The problem

Let us remind ourselves of the discussion on the heat of a candle in the beginning of Chapter 4. We argued that a fundamental solution, i.e. the spreading of heat from a point source, needs to spread out continuously along the geodesic lines (in the case of our world, i.e. $\mathbb{R}^n$, these are straight lines pointing radially outward). Moreover, we argued that as we go further away from the candle, the heat we feel from it should approach 0. This must be true for any distance from the candle. In other words, the felt heat must always be less (or equal in the case of $t = 0$, for which the heat has not spread out yet) for larger distances from the source at any time $t > 0$. Therefore, we must have that the fundamental solution must be decreasing monotonically as a function of distance. By inspecting the fundamental solution on $\mathbb{R}^n$ given by Equation (4.10) (or by simply being aware of the properties of an n-dimensional Gaussian function), we see that it is indeed a monotone decreasing function of $\|y - x\|^2$.
Now, we want to generalize this to Riemannian manifolds. We want to prove the following (assuming a fundamental solution exists for *all* Riemannian manifolds, not only the compact cases).

Conjecture 6.1. Let (M, g) be a Riemannian manifold and let p be the fundamental solution on M with respect to the heat equation with heat operator $L_y := \Delta_g - \partial_t$ operating on the second spatial variable y. Let $d(x, y)$ be the geodesic (or Riemannian) distance defined in Definition 3.52. Then, for any fixed $(x, t) \in M \times (0, \infty)$, the fundamental solution $p(x, y, t)$ is a pointwise monotone decreasing function of $d(x, y)$.

Spoilers: we will not prove this here. In fact, this has (as far as we are aware) never been shown! It turns out that this is very hard to prove and has only been proven in simple spaces. Before discussing the proofs that show this, let us emphasize why this is important.

To see the importance, let us think about what would be the case if Conjecture 6.1 *weren't* true. In analogy with the candle example, that would mean that there

is a pair of distances d_1, d_2 from the candle with $d_2 > d_1$ where a person standing at a distance d_2 would feel more heat from the candle than you, if you are standing at distance d_1. In other words, the heat must have travelled along a non-geodesic path and turn up at d_2 (for instance, it would go upwards a few metres, then go above the other person and travel back down or something similar). This would be absurd and it would contradict the behaviour of heat in general. Intuitively, we know that the "detour" of heat cannot happen. Therefore, it seems excessively logical that Conjecture 6.1 is true.
By this discussion, we see that the theorem is closely related to the statement that **heat always travels along geodesics** for any Riemannian manifold. Of course, intuitively, it must be the case since geodesics are the paths that minimize energy and, as we have observed in nature, processes always take the path of least energy (the second law of thermodynamics, in essence).
Essentially, if this theorem is not true, then "nature" as we know it does not apply to Riemannian manifolds (or actually, on any semi-Riemannian manifold including the mathematical model for describing spacetime). The repercussions of this are tremendous for physics as we know it.

It is very peculiar indeed that such an intuitive and important statement is so hard to rigorously show (or, at least, for mathematicians nowadays). Luckily, there are cases of Riemannian manifolds that do satisfy Conjecture 6.1. This indicates that it must be true for all of them (this is similar to the Riemann hypothesis. A quick reminder: the Riemann hypothesis says that the analytic continuation of the analytic function $f(s) = \sum_{i=0}^{n} \frac{1}{n^s}$, $s \in \mathbb{C}$, that is, the Riemann zeta function, only has roots for $s = -2m$ for $m \in \mathbb{N}$ and $\text{Im}(x) = 0$ or for x so that $\text{Re}(s) = 1/2$. Computational calculations were done on billions and billions of cases and all of them satisfied the hypothesis, but the conjecture has yet to be proven).

6.1.2 Verified cases

In 1981, Cheeger and Yau showed that Conjecture 6.1 holds for the spaces $\mathbb{R}^n$, $\mathbb{H}^n$ and S^n (where $\mathbb{H}^n$ denotes the n-dimensional upper-half plane $\{(x_1, \ldots, x_n) \mid x_n > 0\}$ equipped with a hyperbolic metric) [14]. Their proof was rather involved. It wasn't until three years ago, in 2018, that Alonso-Orán et al. introduced a more elegant proof for the same spaces, but it included some generalizations [15]. The proof made use of the *parabolic maximum principle*, or actually, the maximum principle for parabolic operators. In particular, they showed that Conjecture 6.1 holds (but not for all $x \in M$, as stated) for

- smooth, compact *hypersurfaces of revolution* around the x_n-axis for fixed points x on the intersection of the x_n axis and the hypersurface of revolution. A hypersurface of revolution is essentially the n-dimensional body that is obtained after revolving the $n-1$-dimensional hypersurface (that is, a submanifold of dimension $n-1$ embedded into the manifold) around the x_n axis,
- spherically symmetric manifolds with bounded curvature (around any point x),
- smooth and connected hypersurface of revolution with non-empty boundary, for x in the intersection of the x_n axis and the interior of the connected hypersurface of revolution (so that the fundamental solution either solves the heat equation with Dirichlet or Neumann boundary conditions. The latter implies that the first spatial derivative of p vanishes at the boundary, rather than p itself).

From these, they were able to extract the cases of $\mathbb{H}^n$ and S^n without ever having to state explicit expressions for the fundamental solutions.
Let us look at a non-spherically symmetric case – the flat torus $\mathbb{T}_L^2$ as discussed in Example 4.34 with $n = 2$, but with a width unequal to the length, i.e. with a certain eccentricity $L > 1$. We then have $\mathbb{T}_L^2 = \mathbb{R}^2/(\mathbb{Z} \times L\mathbb{Z})$. We can obtain from the Poisson summation formula that the fundamental solution is given by

$$p_{\mathbb{T}_L^2}(0, y, t) = \frac{1}{4\pi t} \sum_{n,m\in\mathbb{Z}} e^{-\left((n-x)^2+(Lm-y)^2\right)/4t}$$

If we fix y, then this function is decreasing for $0 < x < 1/2$, since then $(n - x)^n$ is increasing for all integers n. Similarly, it is decreasing for fixed x if $0 < y < L/2$. Now, since the radial derivative is a combination of derivatives with respect to x and y, we must have that the function is decreasing in any radial direction. Therefore, it must be decreasing along geodesics at least as y is outside of the cutlocus of x.

This leads us to the following: if we want to find out if the fundamental solution is decreasing as a function of geodesic distance on *all of* M, perhaps we can find a *region* in M that does satisfy the property. Upon inspecting the case of the sphere and the torus, we might suggest that this is always true outside of the collection of cutloci in M. We will not discuss this here in detail, but there is a good chance that this approach might work for all Riemannian manifolds. This is because the sphere is essentially one of the worst cases there is, since, for example, the south pole can get heat from all points through a geodesic on the arc on the great sphere connecting them. However, there is an infinite amount of great spheres connecting the south pole and the north pole. Multiple charts are needed to describe the geodesics along these great spheres.

6.2 Escaping Brownian motion

As we have seen, the fundamental solution on a compact Riemannian manifold integrates to unity, i.e. $\int_M p(x, y, t)dV_g(y) = 1$ for all $(x, t) \in M \times (0, \infty)$. However, for a Riemannian manifold in general, this need not be the case. This has to do with the following notion. Remember that, for the fundamental solution of the heat equation p,

$$\mathbb{P}_x(t < e) = \int_M p(x, y, t)dV_g(y).$$

Definition 6.2. Let (M, g) be a Riemannian manifold. Then M is called **stochastically complete** if for all $x \in M$, a Brownian motion starting from x does not explode with probability 1. That is,

$$\mathbb{P}_x(e = \infty) = 1 \quad \forall x \in M$$

or, equivalently, it means that M is stochastically complete if and only if the fundamental solution integrates to 1.

Now what does it mean for a Riemannian manifold to *not* be stochastically complete? Well, we have actually already seen an example of this. Namely, let $D \subset M$ be any relatively compact domain on M such that $M \backslash \overline{D} \neq \emptyset$. Then

$$\mathbb{P}_x(t < \tau_D) = \int_D p_D(x, y, t)dV_g(y) < 1$$

as we have seen in SECTION 5.3, meaning that D is not stochastically complete. Of course, D is not geodesically complete either (cf. Definition 3.65), since there are curves that have their image outside of D for some $t \in \mathbb{R}$. However, stochastic completeness does not imply geodesic completeness.

Example 6.3. Let $M = \mathbb{R}^2 \backslash \{0\}$. Since $\mathbb{R}^2$ is stochastically complete and a single point does not contribute to an integral (or volume), M is stochastically complete. However, as mentioned while defining the Riemannian distance function (i.e. geodesic distance) in Definition 3.52, we saw that there exists no minimizing geodesic between points $p, q \in M$ such that $p = -q$. Hence, M is not geodesically complete.

Next, we let (M, g) be a stochastically incomplete Riemannian manifold, i.e.

$$\int_M p(x, y, t) dV_g(y) < 1.$$

What does this say about the behaviour of Brownian motion on M? Probabilistically it says that there is a certain moment in time when the Brownian motion stops "running". Formally, it means that there exists a stopping time τ called the *lifetime* of Brownian motion on M, such that Brownian motion X_t "disappears into infinity" as $t \to \tau$, that is,

$$\lim_{t \to \tau} X_t = \infty_M$$

where ∞_M denotes the "point at infinity" belonging to the one-point compactification of M, denoted $\widehat{M} = M \cup \{\infty_M\}$ [12]. So why is it important to study this? Let us give a short, rather heuristic argument in light of physics.
In physics, Brownian motion is often related to the random movement or diffusion of particles or energy (heat). Such particles or energy have a certain "mass" (equivalent to energy). Say, we have found a part of spacetime in the universe that is a non-stochastically complete manifold. Then it means that mass can just "disappear" in that part of spacetime! This can (highly unlikely, of course) be in contradiction with the conservation laws of physics, or with the idea that our universe is completely isolated from other universes (or whatever is outside of our universe).
A set of geometric properties that ensure stochastic incompleteness were derived by Ikeda and Watanabe [16]. The main ingredient of those properties is the *Ricci curvature* of the manifold, denoted $\text{Ric}_M(x)$ which essentially gives a kind of measure as to how much the geometry that is induced by the metric g is different from the Euclidean metric in the point x. They showed the following:

Proposition 6.4. Let (M, g) be a geodesically complete n-dimensional Riemannian manifold with $n > 1$ and fix $o \in M$. Let $r(x) := d(x, o)$ and let $\kappa(r)$ be a nonincreasing continuous map $[0, \infty) \to \mathbb{R}$ such that $\kappa < 0$ for all $r \in \mathbb{R}$, given by

$$\kappa(r) \leq \frac{1}{n-1} \inf\{\text{Ric}_M(x) \mid r(x) = r\}.$$

Then, M is stochastically complete if, for fixed $c > 0$,

$$\int_c^\infty \frac{1}{\sqrt{-\kappa(r)}} dr = \infty.$$

More importantly, under a few minor extra assumptions, Varopoulos [17] showed that if

$$\int_c^\infty \frac{1}{\sqrt{-\kappa(r)}} dr < \infty,$$

then M is *not* stochastically complete. Thus, as long as that integral diverges, we should not have to worry about mass disappearing.
How can we interpret "mass disappearing into infinity"? For illustration, we look at a simple example. Let $H^n = \{(x_0, \dots, x_n) \mid x_0^2 = 1 + x_1^2 + \dots + x_n^2,\ x_0 > 1\}$ be the n-dimensional hyperbolic space (note that here, Index$(g) = 1$ and hence H^n is a semi-Riemannian manifold). It is important to mention that this space is *not* an example of a stochastically incomplete manifold, since it has a constant negative curvature and therefore satisfies the conditions in Proposition 6.4. However, it is a space that allows us to gain insights into what it means for "mass to disappear into infinity".

Definition 6.5. Let $p \in H^n$. Let $S = (-1, 0, \dots, 0)$. Next, denote P the unique intersection of the line through p and S and the hypersurface $x_0 = 0$. Doing this for all $p \in H^n$ we obtain the **n-dimensional Poincaré ball**

$$B_P := \{P \mid p \in H^n\}$$

Let us look at the 1-dimensional Poincaré line (one can verify that this is $(-1, 1)$). This is an example of a stochastically incomplete space (we will not explicitly show this by Varopoulos' criterion). This is a result of the behaviour of the induced metric near ± 1. In order for P to be near to the edge, the point $p \in H^n$ needs to approach infinity in the x_0 direction. Near 0, the distance between two points (say $-a, a, a \ll 1$) is almost equal to their Euclidean distance. However, near say $3/4$, the point p on the hyperbolic plane needs to move exponentially more to reach p_a, for which its image on the Poincaré line $P_a = 3/4 + a$. Hence, the metric is "stretched" more as we P approaches 1. But it can never reach 1, since the metric "diverges" as $P \to 1$. Thus it needs to travel an infinite distance to get to 1.
Thus, imagine if we have a Brownian motion on the Poincaré line, starting at 0. Then, after some time, the Brownian motion can be approaching 1 more and more. In fact, it can be stuck approaching 1 forever, without ever reaching it. This is more or less an interpretation of vanishing Brownian motion.

6.3 Short-time asymptotics

6.3.1 Asymptotic Expansion

As illustrated in Example 4.34, it can be very hard to define an explicit expression for the fundamental solution of the heat equation, even for simple spaces. However, sometimes one wants to study the *behaviour* of heat on a Riemannian manifold rather than just know that a fundamental solution *exists*. Since we can expect heat to travel along geodesics, a logical first step to take is to look at parts of the manifold where geodesics are well-defined and where their length agrees with the Riemannian distance between their endpoints (i.e. where the geodesics are minimizing). In other words, for fixed $x \in M$ where (M, g) is a Riemannian manifold, we look at points y that are *outside of* the *cutlocus* of x (remember from Definition 3.64 that we could interpret the cutlocus of x as the set of points in M for which geodesics stop being minimizing). We denote the set of points that are in the cutlocus of x as C_x. It turns out that the problem is larger for points within each other's cutlocus since there, the fundamental solution of the heat equation depends heavily on the geometric strucure of the minimizing geodesics between them. In particular, it gives a problem because there may be multiple minimizing geodesics between a given pair of points in each other's cutlocus (for example, for a pair of antipodal points on the sphere). As Hsu showed [11], it turns out that fundamental solutions of the heat equation behave very nicely outside of the cutloci on a small time scale, and within the very limit of $t \to 0$ they behave even nicer (with "nice" being "Euclidean-like"). They showed the following.

Theorem 6.6. ***(Asymptotic Expansion).*** *Let (M, g) be an n-dimensional geodesically complete Riemannian manifold. Consider the set of all cutloci,*

$$C_M := \{(x, y) \in M \times M \mid y \in C_x\}.$$

Then, there exists a family of smooth functions $H_i \in C^\infty((M \times M)\backslash C_M)$, $i = 0, 1, 2, \ldots$ such that, on any compact subset $K \subset (M \times M)\backslash C_M$, the asymptotic relation

$$p(x, y, t) \propto \frac{1}{(2\pi t)^{n/2}} e^{-d(x,y)^2/2t} \sum_{i=0}^{\infty} H_i(x, y) t^i \tag{6.1}$$

holds uniformly as $t \to 0$, and such that $H_0(x, y) > 0$ and $H_0(x, x) = 1$. Moreover, we have

$$\lim_{t \to 0} \log\big(p(x, y, t)\big) \cdot t = -\frac{1}{2} d(x, y)^2.$$

The proof makes use of a compact domain D of M and the fact that the fundamental solution on M (denoted p_M as before) has a component p_D for times $t < \tau_D$ and a component that accounts for the expectation value of p_M in the case $t \geq \tau_D$. Then, by the compactness of D it is isomorphic to some compact manifold N, and we can estimate the upper bound for $|p_M - p_N|$ and then let D approach M.
In addition, Hsu showed the identity of H_0. Let $\exp_x : T_x M \to M$ be the Riemannian exponential map at $x \in M$ and let $J(\exp_x)(\xi)$ be the associated Jacobian of $\exp_x$ at $\xi \in T_x M$ (remember that this denotes how much volume elements are "stretched out" or "contracted" in $\exp_x(\xi)$ with respect to the Euclidean volume at ξ). Now, let $y \in M$ and let $\xi = \exp_x^{-1}(y)$, then

$$H_0(x, y) = \frac{1}{\sqrt{J(\exp_x)(\xi)}}.$$

Theorem 6.6 allows us to look at how heat spreads along manifolds on small time scales. For example, on the n-sphere.

Example 6.7. Consider the n-sphere S^n. Remember from Example 3.16 that the standard metric on the 2-sphere is given in spherical coordinates by

$$g_{\text{std}}^{\text{sph}}(\phi, \theta) = \begin{pmatrix} \sin^2\theta & 0 \\ 0 & 1 \end{pmatrix}$$

and that we could also write this as $g_{\text{std}}^{\text{sph}} = \sin^2(\theta)(d\phi \otimes d\phi) + d\theta \otimes d\theta$. Let g_{S^n} denote the standard metric on S^n in *polar* coordinates. Then, we can inductively find that

$$g_{S^n} = dr \otimes dr + \sin^2(r)(d\theta \otimes d\theta),$$

where dr denotes the angle that is introduced when going from S^{n-1} to S^n and where $d\theta \otimes d\theta$ denotes the standard metric on S^{n-1}. We are going to derive the Jacobian of the exponential map $\exp_x$ at y. Let $(e_1, \ldots, e_{n-1})$ be an orthogonal basis of S^{n-1} at the angular position θ and let e_r be the unit vector in the radial direction on $T_x M$

(remember that $T_xS^n \cong \mathbb{R}^n$) based at y. The metric g_{S^n} tells us that if we go from T_xS^n (i.e. from the Euclidean space) to S^n, then the radial unit vector stays the same but the angular basis vectors are contracted by a factor $\sin(r)$, where $r = d(x,y)$. Hence, the Jacobian at y of the exponential map at x is $J(\exp_x)(\exp_x^{-1}(y)) = \left(\frac{\sin(r)}{r}\right)^{n-1}$ and we have

$$H_0(x,y) = \left(\frac{r}{\sin(r)}\right)^{(n-1)/2}.$$

Now, by Theorem 6.6 we have that

$$p_{S^n}(x,y,t) \propto \frac{1}{(2\pi t)^{n/2}} e^{-d(x,y)^2/2t} \left(\frac{d(x,y)}{\sin(d(x,y))}\right)^{(n-1)/2}$$

for all $d(x,y) < \pi$ (this criterion essentially says that x and y cannot be antipodal points so that there is a unique geodesic between them). This expression is a lot better to work with than for example the expression of the fundamental solution of the heat equation on S^1, see Example 4.34.

Moreover, using Theorem 6.6 we can use the fundamental solution to describe Brownian motion on small time scales in a *geodesic ball* around its starting point, that is, a ball on which the Riemannian exponential function is a diffeomorphism on its image (which is again a ball). Remember that the radius of these balls for a point $x \in M$ is given by its injectivity radius, i.e. injrad(x). Hsu showed the following result.

Proposition 6.8. Let $\epsilon := \inf\{\text{injrad}(x) \mid x \in M\}$. Let K be a compact subset of the n-dimensional geodesically complete Riemannian manifold (M,g). Let $r < \epsilon$. There exists a strictly positive map $c_r \in C^\infty(K)$ such that

$$\mathbb{P}_x(t \geq \tau_r) \propto \frac{1}{t^{(n-2)/2}} e^{-r^2/2t} c_r(x)$$

uniformly as $t \to 0$, where τ_r denotes the first exit time of the ball around x with radius r.

Hence, we can approximate the behaviour of Brownian motion in geodesic balls (as soon as it leaves a ball with arbitrary radius r that lies completely within the geodesic ball).

6.3.2 Distant points

Theorem 6.6 describes the behaviour of fundamental solutions to the heat equation when x and y are not in each other's cutlocus. As mentioned, describing its behaviour if $y \in C_x$ is rather complicated due to the dependence on the structure of the set of minimizing geodesics between them. We define this set as $\Omega_{x,y} := \{\text{minimizing geodesics between } x \text{ and } y\}$. In the last section of this thesis, we discuss the behaviour of fundamental solutions of the heat equation on a small time scale when x and y are far apart, perhaps in each other's cutlocus. For example, for antipodal points on a sphere. Throughout this section, we assume that the n-dimensional Riemannian manifold (M,g) is compact and geodesically complete.
The first important result is Varadhan's relation. We have actually already seen this: it is the second implication of Theorem 6.6. However, Varadhan's relation actually implies that this holds for *all* points in a Riemannian manifold, rather than just those outside of each other's cutlocus.

Theorem 6.9. ***(Varadhan's relation).*** *Let $K \subset M \times M$ compact. Let p be the fundamental solution of the heat equation on (M, g). Then, we have uniformly,*

$$\lim_{t \to 0} \log \big(p(x, y, t)\big) \cdot t = -\frac{1}{2} d(x, y)^2$$

for all $(x, y) \in K$.

Varadhan proved this in 1967 [18]. Hsu proved this by proving that it is the upper bound for $\limsup_{t \to 0} \log \big(p(x, y, t)\big) t$ and the lower bound for $\liminf_{t \to 0} \log \big(p(x, y, t)\big) t$ [11]. Thus, we can expand the argument outside of the cutlocus to inside as well, for small times t, any "heat flow" on M is describable using the fundamental solution of the heat equation. Note that since M is compact, we can choose K to be $M \times M$.
Remark. The terms "inside" and "outside" of the cutlocus may be confusing in this context. Remember that the "outside" of the cutlocus is a "nice" place where geodesics are unique and minimizing and where they can be described by the Riemannian exponential function. The "inside" is the space beyond this nice place where other geodesics become minimizing as well, such as antipodal points on a sphere.

What this means for Brownian motion is the following.

Collorary 6.10. Let $D \subset M$ open and connected, with a smooth boundary ∂D. Fix $x \in D$. Then

$$\lim_{t \to 0} t \log \big(\mathbb{P}_x(t > \tau_D)\big) = -\frac{1}{2} d(x, \partial D)^2$$

where $d(x, \partial D) := \inf\{d(x, y) \mid y \in \partial D\}$.

In other words, on a scale of small time, the probability that the Brownian motion will leave the region D is directly dependent on the distance from the boundary of that region. Intuitively, this is logical – but it is indeed reassuring to find that Brownian motion admits such behaviour at least in the limit of small t.

Next, we look at small time behaviour of points specifically inside each other's cutlocus. Let $x, y \in M$ and let $\Omega_{x,y}$ be the set of minimizing geodesics connecting x and y.

Definition 6.11. We define the set that contains the "half point" for all minimizing geodesics in $\Omega_{x,y}$ (i.e. the geodesics evaluated at half the distance between x, y) as

$$\Omega_{x,y}^{1/2} := \Big\{ c\Big(\frac{d(x, y)}{2}\Big) \;\Big|\; c \in \Omega_{x,y} \Big\}.$$

Moreover, the **ϵ-neighborhood of** $\Omega_{x,y}^{1/2}$ is defined as

$$O_\epsilon(x, y) = \Big\{ z \in M \mid d\big(z, \Omega_{x,y}^{1/2}\big) < \epsilon \Big\}.$$

Obviously, since geodesics are not minimizing in the cutlocus, we have that $\Omega_{x,y}^{1/2} \cap (C_x \cup C_y) = \emptyset$. Hence $d\big(\Omega_{x,y}^{1/2}, (C_x \cup C_y)\big) > 0$ and thus we can choose $O_\epsilon(x, y)$ to be disjoint of the cutloci as well (for small enough ϵ).
Hsu showed the following important relation.

Theorem 6.12. *Let p be the fundamental solution of the heat equation on (M, g). Let H_i be as in Theorem 6.6. Then there exists a constant $\lambda > 0$ so that for every $(x, y, t) \in M \times M \times (0, 1)$*

we have

$$p(x,y,t) = \big(1 + f(x,y,t)\big)\frac{e^{-d(x,y)^2/2t}}{(\pi t)^n} I_\epsilon(x,y) \tag{6.2}$$

where f is a function so that $|f(x,y,t)| \leq \lambda t$ and where the integral $I_\epsilon(x,y)$ is given by

$$I_\epsilon(x,y) = \int_{O_\epsilon} H_0(x,z)H_0(z,y)e^{-Q_{x,y}(z)/2t} dV_g(z)$$

in which we have used the expression $Q_{x,y}(z) := 2d(x,z)^2 + 2d(z,y)^2 - d(x,y)^2$.

This result allows us to calculate the short-time asymptotic behaviour of the fundamental solution to the heat equation on manifolds even for points in each other's cutlocus. Let us conclude this chapter by illustrating this in a specific case – antipodal points on the n-sphere.

Example 6.13. Let S^n be the n-sphere. Let $N = (1, 0, \dots, 0)$ be the north pole and let $S = (-1, 0, \dots, 0)$ be the south pole. Let p be the fundamental solution of the heat equation on S^n with the metric in polar coordinates as in Example 6.7. Note that $\Omega^{1/2}_{N,S}$ in this case is the great sphere S^{n-1} perpendicular to all the geodesics between N and S, i.e. $\{(x_0, \dots, x_{n-1}) \mid x_0 = 0\}$ (the geodesics themselves are all semi great spheres). For all $z \in S^n$, let $r := d(N,z)$. Then, in other words, we have that $\Omega^{1/2}_{x,y}$ is given by $r = \pi/2$. We have $d(S,z) = \pi - r$ (since there is always a great sphere through z, N and S). Then, we have

$$Q_{N,S}(z) = 2r^2 + 2(\pi - r)^2 - \pi^2 = 4\Big(r - \frac{\pi}{2}\Big)^2.$$

Remember from Example 6.7 that we have found for all $x, y \in M$ with $d(x,y) < \pi$

$$H_0(x,y) = \Big(\frac{d(x,y)}{\sin(d(x,y))}\Big)^{(n-1)/2}$$

and that the volume element on the sphere (i.e. the Jacobian) is $\sin^{n-1}(r)drd\theta$. Then, by Theorem 6.12 we have, again using the polar coordinates integration technique, that

$$p(N,S,t) \propto \frac{e^{-\pi^2/2t}}{(\pi t)^n} \int_{|r-\pi/2|<\epsilon} e^{-2(r-\pi/2)^2/t} h(r) dr \int_{S^{n-1}} d\theta$$

where the auxiliary radial function h is given by

$$h(r) = \left(\frac{r}{\sin(r)} \cdot \frac{\pi - r}{\sin(\pi - r)}\right)^{(n-1)/2} \sin^{n-1}(r).$$

Now, we can apply the Laplace approximation to see that

$$\int_{|r-\pi/2|<\epsilon} e^{-2(r-\pi/2)^2/t} h(r) dr \propto \sqrt{\frac{\pi t}{2}} h(\pi/2)$$

and hence we find the desired result

$$p(N,S,t) \propto \frac{e^{-\pi^2/2t}}{\sqrt{2\pi} 2^{n-1}} \cdot \frac{1}{t^{(2n-1)/2}} \cdot \mathrm{vol}(S^{n-1}) \tag{6.3}$$

where we have $\mathrm{vol}(S^{n-1}) = \int_{S^{n-1}} d\theta$. For literature on the Laplace approximation, see Copson [19].

Hence we have found how to describe the fundamental solution to the heat equation and Brownian motion in the limit of short time in points of compact Riemannian manifolds inside and outside of cutloci.

7. Conclusion

In this thesis, we have discussed the following.
In order to look at the heat equation on Riemannian manifolds, we needed some basics on differential geometry. In Chapter 2 we have seen what it means for functions from and to manifolds to be differentiable. This gave rise to differentiable manifolds, and what their derivatives are, i.e. differentials, using tangent vectors in tangent spaces. Tangent spaces can be identified with Euclidean space through isomorphism (the differential of charts). We saw that we could use derivations to express tangent vectors in terms of the basis of the set of all derivations, i.e. the directional derivative in the direction of the components of charts. We explored how derivatives depend on the choice of charts. Finally, we defined vector fields on differentiable manifolds using the tangent bundle and discussed what it means for vector fields to be differentiable.

After this, we dived into Riemannian geometry theory in Chapter 3. We saw that Riemannian manifolds are differentiable manifolds equipped with a (Riemannian) metric that assigns to each point in the manifold a non-degenerate, positive definite, symmetric bilinear form acting on the tangent space of that point. We introduced the cotangent space while determining the dependence of the metric on the choice of charts and saw some examples of Riemannian manifolds. Next, we defined the derivative of vector fields and along the way introduced the Levi-Civita connection and saw that is unique for each point on a Riemannian manifold and that it can be expressed in terms of Christoffel symbols. Moreover, vector fields could be differentiated in any direction using covariant derivatives, or along curves. This gave rise to the idea of geodesics, which are curves that have a vanishing temporal derivative of the velocity field. We saw that we could interpret this as the fact that geodesics that feel no acceleration/force and hence it minimizes the energy of of all curves between two points in a variation. In addition, the energy is constant and so is the metric in the directional derivative along the geodesic in any point on the curve. We defined the Riemannian exponential map and the injectivity radius and showed that it is a diffeomorphism on balls with a radius equal to the injectivity radius and that the map constitutes the unique geodesic on such a ball. We saw that doing geometry was easier using geodesic normal coordinates. Finally, we introduced the Riemannian distance function and lengths of curves and noted that the Riemannian metric induces the original metric and topology on manifolds. We saw what it means for geodesics to be minimal and introduced the important notions of a cutlocus and a geodesically complete manifold.

Next, we discussed the heat equation. The main objective of Chapter 4 was to show the existence and uniqueness of the fundamental solution of the heat equation, that is, the solution that starts as a point source with temperature 1, on compact Riemannian manifolds without boundary. First, we introduced the (in)homogeneous heat equation and the heat operator. Next, we defined the Laplace-Beltrami operator on Riemannian manifolds using the gradient and divergence function induced by the Riemannian metric. We then introduced the heat equation on Riemannian manifolds and defined integration on them. We proved that solutions to the initial conditions version of the heat equation were unique and using Duhamel's principle we found that the fundamental solution is unique (if it exists). Subsequently, we took a detour to the finding of the fundamental solution on $\mathbb{R}^n$ using Fourier analysis and it yielded a good ansatz for the fundamental solution on compact Riemannian manifolds. However, this did not solve the heat equation and we modified the expression to decay to zero as the time approaches 0 when the heat operator acts on it. Then, we introduced a bump function to extend the argument beyond injectivity radii. Finally, we used a series of convolutions to find an expression for the fundamental solution of the heat equation on Riemannian manifolds, showing its existence. We concluded with an example: the torus.

In Chapter 5, we discussed the stochastic processes known as Brownian motion on Riemannian manifolds. The main objective here was to show that the transition density, i.e. the probability that the Brownian motion will be in a certain region of the manifold after a certain time, is defined by the fundamental solution of the heat equation. We defined Brownian motion as a martingale where each random variable has a Gaussian distribution with the variance equal to the time during which the Brownian motion elapses. We looked into the notions of exit time and explosion time. We took yet another detour to the transition density function on $\mathbb{R}^n$ and verified that the main assertion of this chapter was satisfied in the Euclidean space. Finally, we used Ito's Lemma to extend this theorem to relatively compact domains and concluded that the density transition function is indeed given by the fundamental solution of the heat equation on compact Riemannian manifolds by letting that domain approach the manifold itself.

In Chapter 6, we discussed examples of the heat equation and Brownian motion on Riemannian manifolds. First, we discussed the intuitive but unsolved statement that the fundamental solution is a decreasing function of Riemannian distance and stressed the importance of this statement. We explored some proofs on specific cases to help illustrate it.
Next, we discussed the notion of stochastically completeness of Riemannian manifolds and the phenomenon of disappearing Brownian motion. We provided some geometric criteria for stochastically (in)completeness and looked into the example of Poincaré balls.
We concluded the body part with a discussion on short-time asymptotics, which provided us with mathematical apparatus to describe complex fundamental solutions with less complex expressions in the limit of small time. We saw that this is a lot harder for points inside of each other's cutlocus and provided an expression of proportionality for the fundamental solution on the n-sphere for points outside of each other's cutlocus. Finally, we saw that approximations can be made for distant points, i.e. in each other's cutlocus as well, such as antipodal points on the sphere.

All in all, we have developed a theoretical framework to show that the heat equation has exactly one fundamental solution on compact Riemannian manifolds and that this fundamental solution is the transition density function of Brownian motion on compact Riemannian manifolds. Future research may be done on proving Conjecture 6.1, developing a system so that the property of stochastically incompleteness can more easily be identified, as well as non-asymptotic behaviour of the heat equation and Brownian motion on Riemannian manifolds. One way to expand the research done in this thesis is to prove the existence of, and to study, the fundamental solution of the heat equation on non-compact Riemannian manifolds, or on semi-Riemannian manifolds (in order to, for example, look at heat or Brownian motion in spacetime bent by gravity).

A. Bibliography

[1] Marcel Berger. "Riemannian Geometry During the Second Half of the Twentieth Century". In: *University Lecture Series, 17, Rhode Island: American Chemical Society* (2000).

[2] Christian Bär. "Differential Geometry Lecture Notes". In: *Universität Potsdam* (Aug. 2013).

[3] Mikio Nakahara. *Geometry, Topology and Physics, 2nd edition*. CRC press, 2003.

[4] Alexander Grigoryan. *Heat Kernel and Analysis on Manifolds*. University of Bielefield, 2009.

[5] Isaac Chavel. *Eigenvalues in Riemannian Geometry*. City University of New York, 1984.

[6] Yaiza Canzani. "Analysis on Manifolds via the Laplacian". In: (2013).

[7] Ngô Quoc Anh. "Laplace-Beltrami operator, read on 02-06-2021". In: (2009). URL: `\url{https://anhngq.wordpress.com/2009/12/05/r-g-laplace-beltrami-operator/}`.

[8] Ngô Quoc Anh. "A Useful Formula, read on 02-06-2021". In: (2009). URL: `\url{https://anhngq.wordpress.com/2009/12/02/r-g-a-useful-formula/}`.

[9] Steven Lalley. *Brownian Motion Lecture 5: University of Chicago*.

[10] Julio D. Rossi Pablo Groisman. "Explosion time in stochastic differential equations with small diffusion". In: (2007).

[11] Elton P. Hsu. *Stochastic Analysis on Manifolds*. American Mathematical Society, 2002.

[12] Elton P. Hsu. "A brief introduction to Brownian Motion on a Riemannian manifold". In: (2002).

[13] Wioletta M. Ruszel. "Stochastic Differential Equations Lecture Notes". In: *Utrecht University, TU Delft* (2019).

[14] S. T. Yau. J. Cheeger. "A lower bound for the heat kernel". In: *Comm. Pur. Appl. Math 34* (1981), pp. 465–480.

[15] D. Martinéz A. Mas D. Alonso-Orán F. Chamizo. "Pointwise monotonicity of heat kernels". In: (2018).

[16] S. Watanabe N. Ikeda. *Stochastic Differential Equations and Diffusion, volume 24*. North Holland, 1992.

[17] N. Varopoulos. "Potential theory and diffusion on Riemannian manifold". In: *Conference on Harmonic Analysis in Honor of Antoni Zygmund, Vol. II, Wadsworth Math. Ser.* (1983), pp. 821–837.

[18] S. Varadhan. "Diffusion processes in a small time interval". In: *Comm. Pure Appl. Math., 20* (1967), 659–685.

[19] E. T. Copson. *Asymptotic Expansions*. Cambride University Press, 1965.