Master Thesis
year 2014-2015

The Laplace Operator on a Riemannian Manifold

Author:
Yoann Trellu
Supervisor:
Prof. Tom Ilmanen

September 3, 2015

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Contents

1 Introduction

The study of diffusion processes on a manifold, very much interesting by itself, also has a practical motivation. We refer to the beginning of Stéphane Lafon's Thesis [1] which gives a good overview of the particular interests of this topic. Two main ideas are:

1. The geometry of a data set in high dimensional space can sometimes be approximated by an underlying submanifold.

2. Diffusion processes explain much of the local geometry on a manifold. As such, they become good candidates to design mappings from this high dimensional space to a low dimensional space while preserving the local geometry.

While his work also includes results about discrete approximations via graphs, we choose to focus on the differential geometry side and display results on the geometrical insight given by diffusion processes on a manifold. This practical application will be the red thread along this work and should explain the choice of results presented.

We first display a few applications which should give an idea on the range of these studies. We then look into functional analysis to derive sufficient conditions on the use of eigen-decomposition of an operator. We finally investigate Sobolev spaces on a manifold to conclude our eigen-decomposition.

The reader unfamiliar with Differential Geometry up to *Stokes Theorem* should refer to the appendices prior to the main part. Functional Analysis up to Riesz Theorem is also assumed.

2 Applications & Motivation

We first look at two simple examples on the use of spectral theory for data compression. We then talk about Principal Component Analysis (PCA), which can be viewed as a linear approximation of the data set geometry. We then extend these approximations to the non linear case. Finally, we present the Elastic Maps algorithm [3] which deals with the non linear representation of data sets. An example is described using a three dimensional representation of shares from the New York Stock Exchange.

2.1 Audio & Image Compression

Audio Compression namely the MP3 format. $\mathcal{M}$ becomes a closed interval $[a, b]$ of the real line.

Image Compression namely the JPEG format. $\mathcal{M}$ becomes a closed subset $[a, b] \times [c, d]$ of $\mathbb{R}^2$.

Both use a spectral decomposition, to keep the principal information by selecting the relevant frequencies.

2.2 Principal Curves and Surfaces

We present here this theory first presented by Trevor Hastie [2]. Let us say we observe a set of points $\{x_1, \cdots, x_n\} \subseteq \mathbb{R}^d$ with an additional noise component $(\epsilon_i)_{i=1,\cdots,n}$

$$x_i = u_i + \epsilon_i.$$

We assume that these observations come from a linear structure, namely that $u_i = At_i$ with $A \in \mathbb{R}^{d\times n}$. Additionally, we assume that the errors $\epsilon \in \mathbb{R}^n$ are distributed such that $\mathbb{E}[\epsilon] = 0$, $\mathbb{E}[\epsilon^T \epsilon] = \sigma^2 I$ for some $\sigma > 0$. We then know from the Gauss-Markov theorem that minimizing

$$D(t, A) := \sum_{i=1}^{n} ||x_i - At_i||^2 \tag{1}$$

leads to the statistical estimate with minimal variance among the unbiased candidates. Since t is obtained by projecting the observations onto A, equation (1) reduces to

$$D(t, A) = D(A) := \sum_{i=1}^{n} ||x_i - AA^T x_i||^2.$$

This minimization problem is solved by taking the eigenvectors from the matrix $X^T X$ where $X := (x_1 \cdots x_n)^T$. Those eigenvectors, once ordered by decreasing order with respect to their eigenvalues, are called *principal components* and form the basis for PCA. Another way to look at PCA is to say that it tries to explain the most variance using suitable linear spaces.

We would like now to get rid of the linearity assumption governing the true underlying $u \in \mathbb{R}^{d \times n}$. Namely, we assume the model

$$x_i = f(t_i) + \epsilon_i. \tag{2}$$

Notice how the linear formulation is a special case of (2). Let us then see how we can generalize the reasoning we held above to extend the first principal component of PCA to a *principal curve.* First, we assume a random vector X in $\mathbb{R}^d$ with continuous probability density $h(x)$. We further define the set $\mathcal{G}$ of differentiable curves parametrized by a real λ in a closed interval, in $\mathbb{R}^d$. In addition, these curves should not intersect, form loops or be tangent to themselves.

Definition (Principal Curve [2])**.** *An element f of $\mathcal{G}$ is a principal curve if it is* self-consistent *for h:*

$$\mathbb{E}_h[X : \lambda_f(X) = \lambda] = f(\lambda)$$

where $\lambda_f(X) := \sup\{\lambda : ||X - f(\lambda)||^2 = \inf_\tau ||X - f(\tau)||^2\}$.

$\lambda_f(X)$ should be interpreted as being the coordinate on the curve being the closest to X, if there are many candidates, the smallest is taken. For example, if f is a a semi-circle with X at its radius center, we take one of the extremities of f depending on the direction taken by the parametrisation by λ. To repeat with words the above definition, each points on the principal curve is the *conditional mean* of all points (coming form h) that project there.

We now look at another property of principal curves. Firstly, let $d(x, f)$ denote the euclidean distance from a point x to its projection $f(\lambda_f(x))$,

$$d(x, f) := ||x - f(\lambda_f(x))||.$$

Secondly, define $D^2 : \mathcal{G} \to \mathbb{R}$ by

$$D^2(f) := \mathbb{E}_h[d(X, f)^2].$$

this allow us to present the following definition.

Definition (Critical Values [2])**.** *If f and g belong to $\mathcal{G}$ and we form $f_\epsilon = f + \epsilon g$, we define f to be a critical value of D^2 if and only if*

$$\frac{dD^2(f_\epsilon)}{d\epsilon}|_{\epsilon=0} = 0.$$

Hence critical values are minimas, maximas or saddle points in the variational sense of the distance function for functions in the class $\mathcal{G}$.

Theorem 1 (Stationarity Property [2])**.** *The curve f is a principal curve of h if and only if f is a critical value of the distance function in the class $\mathcal{G}$.*

Now if we define by $\mathcal{L} \subseteq \mathcal{G}$ the class of straight lines, we can prove that principal components are the critical values of the distance function within the class $\mathcal{L}$. As such, this result tells us that principal curves are indeed a generalization of principal components.

Principal surfaces come as well from a generalization of the linear space spanned by the first two principal components. Its definition will closely follow the one for principal curves.

There are further results that could be discussed such as existence and bias matters. We refer the reader to [2] for further information.

2.3 Elastic Nets

In practice we are not given a continuous probability distribution h on $\mathbb{R}^d$ but a dataset $D \subseteq \mathbb{R}^d$. One therefore assumes that D is subordinate to some h, an unknown continuous probability distribution. Two solutions exists to estimate h. A first one described in [2], uses *kernel smoothers* to estimate the expectation seen in definition (2.2).

A second solution thoroughly explained in [3], uses a minimization procedure over a set of nodes $W_j \in \mathbb{R}^d, j \in \{1, \cdots, k\} := N$, edges $(i, j) \in E \subseteq N \times N$ and *bending ribs* $(i, j, l) \in G \subseteq E \times E$. G has to be a subset of pairs of edges sharing a common node. They will act as a discrete approximation for the underlying principal object. In addition, we consider this first term

$$E_a := \sum_{j=1}^{k} \sum_{s \in K_j} ||s - W_j||^2$$

with $\{K_j, j \in \{1, \cdots, k\}\}$ a partition of our data set D. Using squared euclidean distances, E_a estimates how close are the nodes W_j to the dataset D. E_a is called approximation energy.

We define a second term

$$E_s := \lambda \sum_{(i,j) \in E} ||W_i - W_j||^2$$

called the stretching energy. This term constrains the set of nodes K to be as uniformly positioned as possible with respect to the edges E.

And finally a third term

$$E_b := \mu \sum_{(i,j,l) \in G} ||W_i - 2W_j + W_l||^2$$

named the bending energy. Notice the discretized second derivative within the norm. This term will penalize (adjusted by μ) the non-linearity of the graph structure formed by (K, E, G).

An optimization procedure called *principal graph strategy* is presented below ([3], pp 109).

1. Make a grid from a number of unconnected nodes (sets of edges and ribs are empty at this stage). Optimize the node positions (i.e., do K-means clustering for large k). The number of nodes k is chosen to be a certain proportion of the number of data points. In our experiments we used k = 5% of the total number of data points.

2. For every node of the grid in position y^i, the local first principal direction is calculated. By local we mean that the principal direction is calculated inside the cluster of data points corresponding to the node i. Then this node is substituted by two new nodes in positions $y^{new1} = y^i + \alpha sn, y^{new2} = y^i - \alpha sn$, where n is the unit vector in the principal direction, s is the standard deviation of data points belonging to the node i, α is a parameter, which can be taken to be 1. An edge is generated between the two new nodes.

3. A collection of edges and ribs is generated, following this simple rule: every node is connected to the node which is closest to this node but not already connected at step 2, and every such connection generates two ribs, consisting of a new edge and one of the edges made at step 2.

4. The node positions are optimized according to the quadratic functional

$$U := E_a + E_s + E_b.$$

Its motivation, further optimization strategies as well as its computational cost can be found in [3].

2.4 General formulation

Our problem can be described as follows:

Given a compact n smooth manifold $\mathcal{M}$ we want to process or find a (set of) function(s) $f : \mathcal{M} \mapsto \mathbb{R}$ satisfying either a physical interpretation or an optimization problem. Notice that for dimensionality reduction, we could be interested by a set of functions $f_i : \mathcal{M} \mapsto \mathbb{R}$, with $i \in \{1, \cdots, m\}$ where typically $m << n$.

3 Basics of Operator Theory

In this section, we consider a general linear operator $A : X \mapsto Y$ with X, Y two suitable function spaces. Our goal will be to understand its behavior with respect to the structure of X and Y, and ultimately come up with an educated choice of function spaces which will have the required properties for the next sections.

Among these properties we can already name the ability to develop spectral theory, which will be useful as we have seen in the applications. We will also take a look at convergence issues - we need to understand in which sense a sequence f_n converges to f in X - as well as smoothness and potential approximations. We assume basic knowledge of functional analysis as seen in appendix D from Evans [8]. Furthermore,

- we use the real field for our normed vector spaces whenever it is not mentioned,

- functions are real-valued.

3.1 Preliminaries

Here are some motivations to investigate the spectral theory surrounding the Laplace operator.

1. A specific class of linear operators will define a basis on our functional space. A basis intuitively brings new lenses with which one can study the functional space. Using a new basis to express a specific function will tell us how this function relates to the underlying operator. The common example concerns the Laplace operator together with its *frequency lens* known as Fourier Analysis.

2. Following the previous item, a function that specifically provides information on its domain can then be analyzed through the above *lens* to give away specific information on the domain.

3. Once the basis is found, the operator can be trivially manipulated which results in easier computations. Additionally, this brings a tool to classify a specific class of linear operators by comparing them via their spectrum. This will essentially be a generalized version of saying that two matrices act in the same way yet using different basis and different points of view.

Now to understand how this specific class of operator is defined, we recall a first lemma and theorem on matrices from linear algebra, which will also make the above ideas rigorous in the finite dimensional case:

Lemma 1 ([7])**.** *Let A be an operator on a finite dimensional inner product space. Then A is normal if and only if the space is spanned by an orthonormal set of eigenvectors of A.*

If we are thus to be motivated by the three previous observations, we need to restrict our attention to normal operators. We will actually restrict ourselves even further, by only considering self-adjoint operators.

Theorem 2 (Spectral Theorem - Finite Dimensions)**.** *Let V be a (n)-finite dimensional inner product space and A a symmetric linear operator on V. Then there exists a finite sequence $(\lambda_1, e_1), \cdots, (\lambda_n, e_n) \in \mathbb{R} \times V$ such that*

1. *$\{e_1, \cdots, e_n\}$ are orthonormal eigenvectors for A and spans the space V.*
2. $$Ax \;=\; \lambda_1 P_1(x) + \cdots + \lambda_n P_n(x) \;=\; \sum_{i=1}^{n} \lambda_i \langle x, e_i \rangle e_i$$

   for all $x \in V$ where P_i denotes the projection along vector e_i.

The above theorem expresses the observations in a rigorous way. Next, let us extend the discussion beyond finite rank operators.

3.2 Compact Operators

Lemma 2 ([7])**.** *A continuous finite rank operator between two normed vector space ($A : X \mapsto Y$) is compact.*

Proof. Let us take a bounded sequence $\{x_n\}_{n\geq 1} \subset X$. Since $(R(A), ||\cdot||_Y)$ is a normed vector space of finite dimension (say n), it can be identified with the space $\mathbb{R}^n$ together with some norm $||\cdot||_{\mathbb{R}^n}$. Now since in finite dimensions all norms are equivalent, we now that a result about convergence holds for the norm $||\cdot||_2$ equivalently holds for the norm $||\cdot||_{\mathbb{R}^n}$. Now since the sequence $\{A(x_n)\}_{n\geq 1}$ is bounded, by continuity of A, it holds by a standard theorem of analysis that it has a convergent subsequence. □

Theorem 3 ([7])**.** *Let X and Y be normed vector space with Y complete. For a sequence $\{T_n\}_{n\geq 1}$ of compact operators and a continuous operator T such that*

$$\lim_{n\to\infty} ||T_n - T|| = 0$$

we have that T is compact.

sketch of proof.
Recall that to show compactness of an operator T, it is sufficient to prove that every bounded sequence $\{x_m\}_{m\geq 1} \subset X$ has a subsequence $\{x_{m_n}\}_{n\geq 1}$ such that $\{Tx_{m_n}\}_{n\geq 1}$ converges in Y. The bulk of the work concerns the building of this subsequence. A *diagonal argument* on the sequences $\{\{x_{m,n}\}_{n\geq 1}\}_{m\geq 1}$ converging respectively for all T_m, shows that there exists a Cauchy subsequence. The completeness of Y is then used to conclude. □

One can go further than lemma (2) to understand how compactness extends continuous finite rank operators. If Y is a Hilbert space, then every compact operator is the limit of finite rank operators. In other words and using theorem (3), for Y a Hilbert space, compact operators are the closure of finite rank operators in the operator norm (or equivalently in the uniform topology).

Theorem 4 (Spectral Theorem - Compact Operators [7])**.** *Let H be a Hilbert space of infinite dimension and A a compact, symmetric operator on H. We further assume that the kernel of H is separable. Then there exists an infinite (countable) sequence $(\lambda_1, e_1), \cdots, (\lambda_i, e_i), \cdots \in \mathbb{R} \times H$ such that the following holds:*

1. *The sequence $\{e_i\}_{i\geq 1}$ is orthonormal and $H = \overline{span\{e_i : i \in \mathbb{N}\}}$.*

2. *The sequence $\{\lambda_i\}_{i\geq 1}$ satisfies $\lim_{n\to\infty} \lambda_n = 0$.*

3. $$Ax = \lim_{n\to\infty} \sum_{i=1}^{n} \lambda_i \langle x, e_i\rangle e_i$$

   for all $x \in H$.

3.3 Semi-Bounded Symmetric Operators

In this subsection we present a few results that will allow us to make use of the previous spectral theorem for a more general class of operators. Indeed, we will see that by inverting *semi-bounded* operators on carefully chosen spaces, we can still get a sequence of eigenvectors related to them.

3.3.1 Dense Domains

Before dwelling into the subject, let us first talk about domains of operators. We previously considered only continuous operators which had the whole Hilbert space as domain. We next see why this can no longer be assumed. First, we need a result that links closeness with boundedness.

Theorem 5 (Closed Graph Theorem [7]). *Let X and Y be Banach spaces and let $T : X \to Y$ be a linear operator. T is bounded if and only if the graph of T $G(T)$ is closed in $X \times Y$.*

Next, we link symmetry and closedness.

Proposition 1. *Let T be a symmetric everywhere defined operator. Then T is closed.*

Proof. Recall that T being closed means that the associated graph $\mathcal{G}(T)$ is closed in $H \times H$. Therefore, let us take a sequence (x_n, Tx_n) in $\mathcal{G}(T)$ converging to (x, y) and show that the limit is still in $\mathcal{G}(T)$. Using the continuity of the scalar product we write

$$\langle Tu, x\rangle = \lim_{n\to\infty} \langle Tu, x_n\rangle = \lim_{n\to\infty} \langle u, Tx_n\rangle = \langle u, y\rangle$$

which holds for all $u \in H$. By symmetry, $Tx = y$ which shows that $(x, y) \in \mathcal{G}(T)$. □

We thus need smaller domains than the whole space H to make sense of unbounded symmetric operators.

Proposition 2 (Dense Domains). *for a subspace $D \subseteq H$, either D is dense in H or there exists a non-trivial subspace $D^\perp$ of H orthogonal to D.*

Proof. D is dense in H if $\overline{D} = H$. Thus it is sufficient to prove that for any subspace D of H, we have $D^\perp = \overline{D}^\perp$.

$$x \in D^\perp \Leftrightarrow \forall\, y \in D,\ \langle x, y\rangle = 0 \Leftrightarrow \forall\, y \in \overline{D},\ \lim_{n\to\infty} \langle y_n, x\rangle = 0 \Leftrightarrow x \in \overline{D}^\perp$$

where $(y_n)_{n\in\mathbb{N}}$ is a sequence in D converging to y (it exists by density) and the continuity of the scalar product is used. □

Now, the existence of a non trivial orthogonal complement to the domain contradicts the existence of a set of eigenvectors spanning the whole space H. Therefore following the two previous propositions, it will be assumed that not necessarily bounded symmetric operators A have dense domains $D(A) \subseteq H$.

3.3.2 Semi-Boundedness

Semi-boundedness for a symmetric operator A on a Hilbert space H means that there exists a constant C such that

$$\langle Au, u\rangle \geq C||u||^2, \,\forall\, u \in D(A).$$

Notice that when $C > 0$ and by applying Cauchy-Schwarz, we have

$$||Au|| \geq C||u||, \,\forall\, u \in D(A). \tag{3}$$

This gives us the existence of its inverse A^{-1} by the injectivity derived from (3). Indeed, the inequality constraints the kernel of A to be the null element. Further, by composing by A^{-1} we have $||A(A^{-1}v)|| \geq C||A^{-1}v||$ which is rewritten as

$$||A^{-1}v|| \leq \frac{1}{C}||v||, \,\forall\, v \in D(A^{-1}) := R(A) \tag{4}$$

the definition of boundedness.

The computations above required the semi-boundedness constant C to be strictly positive. If C is negative, this can be forced by adding to A a constant term:

$$\langle (A+(|C|+1)I_d)u, u\rangle = \langle Au, u\rangle+(|C|+1)||u||^2 \geq -|C|||u||^2+(|C|+1)||u||^2 = ||u||^2$$

for all $u \in D(A)$. It can be proven that this modification does not invalidate the reasoning that will follow. We assume from now on that C is strictly positive.

3.3.3 Self-Adjoint Operators

Having a symmetric semi-bounded operator

$$A : D(A) \subseteq H \mapsto H$$

we would like to associate to it a sequence of eigenvectors and eigenvalues as we have seen previously for compact operators. Since compactness requires boundedness, the first little trick uses the inverse which was shown to exists and is bounded. Further, if this inverse fullfills the requirements for Theorem (4), we then have

$$A^{-1}u = \lambda u \;\Leftrightarrow\; A(A^{-1}u) = \lambda Au \;\Leftrightarrow\; Au = \frac{1}{\lambda}u$$

for all eigen-pairs (u, λ) of A^{-1}. Therefore we can associate to A a sequence of pairs $(u_i, \frac{1}{\lambda_i})$ such that $\{u_i\}$ spans H and $\lim_{i\to\infty} 1/\lambda_i \to \infty$. Notice that writing A as a linear combination of eigenvectors is delicate; A being unbounded we cannot make sense of the limit in point 3 of Theorem (4). Now to apply the spectral theorem for compact operators on A^{-1}, the inverse has assumptions to fulfill.

Everywhere defined. As previously said, the domain cannot allow for a non-trivial complement. The domain has thus to be at least dense. In fact it also needs to be complete since the proof of the theorem uses the completeness of H. Yet a complete subspace of a metric space is closed hence the everywhere defined requirement. This point will be the object of much of the work in this subsection.

Symmetry. The symmetry follows easily by composing with the operator A.

Compactness. As shown before, we already have the boundedness which implies the continuity for our inverse. It would then be sufficient to show that the image of our inverse is compactly contained in H. Indeed, the composition of a continuous and compact operator is compact. Within this section we will not prove the existence of such a compact embedding of spaces. This will be done in the particular case of the Laplace operator. We will however hint at such a space deriving the necessary continuous embedding property.

Let us take a look at the first requirement. We write the inverse of A as

$$A^{-1} : R(A) \subseteq H \mapsto D(A) \subseteq H.$$

The goal is to make sure that $R(A)$ is closed to have the everywhere defined assumption. A nice tool will help us in that matter.

Definition (Adjoint)**.** *The Adjoint $T^\star$ of an operator T has the following domain*

$$\varphi \in D(T^\star) \;\Leftrightarrow\; \exists\, \eta \in H \;\; s.t. \;\; \langle T\psi, \varphi\rangle = \langle \psi, \eta\rangle \;\forall\, \psi \in D(T)$$

and is defined with

$$T^\star\varphi = \eta \Leftrightarrow \langle T\psi, \varphi\rangle = \langle \psi, \eta\rangle \;\forall\, \psi \in D(T).$$

We make a few observations:

- The Riesz theorem and the fact that $D(T)$ is dense ensures that the adjoint is well-defined, namely the unicity of η.
- For a symmetric operator T, its adjoint coincides with T on $D(T)$. Further, $D(T) \subseteq D(T^\star)$ and $R(T) \subseteq R(T^\star)$.
- For a symmetric operator T, if $D(T) = D(T^\star)$ then T is called *self-adjoint*.

To further understand self-adjointness, let us look at their graphs.

Proposition 3 (Characterization of Self-Adjointness [4])**.** *For a symmetric operator T, its graph is related to its adjoint as follows:*

$$\mathcal{G}(T^\star) = V(\mathcal{G}(T))^\perp$$

with $V(x,y) = (y,-x)$ for all pairs $(x,y) \in H \times H$.

Proof. We first rewrite the self adjointness property:

$$(x,y) \in \mathcal{G}(T^\star) \;\Leftrightarrow\; x \in D(T^\star),\, y = T^\star x \;\Leftrightarrow\; \forall u \in D(T),\; \langle Tu, x\rangle = \langle u, y\rangle. \tag{5}$$

Next, for any pair (x,y) in the orthogonal complement of $V(\mathcal{G}(T))$ we have

$$0 = \langle (x,y), (Tu,-u)\rangle_{H\times H} := \langle x, Tu\rangle + \langle y, -u\rangle = \langle Tu, x\rangle - \langle u, y\rangle$$

for all $u \in D(T)$. This completes the proof. $\square$

Notice that the operator V acting on the "plan" $H \times H$ shifts the graph by an orthogonal angle to the right. Thus the right hand side can be thought of as taking the complements twice to get back on our feet. Using the proof of proposition (2) and the continuity of V, we may additionally write

$$\mathcal{G}(T^\star) = V(\mathcal{G}(T))^\perp = \overline{V(\mathcal{G}(T))}^\perp = V(\overline{\mathcal{G}(T)})^\perp. \tag{6}$$

This should help understand from a different perspective, how the self-adjoint extension *completes* a plain symmetric operator.

Notice how the above proposition characterizes self-adjointness, as the graph of an operator is uniquely determined. We next continue with our understanding of this property, namely how it behaves under the inverse operation.

Proposition 4 (Inverse of a self-adjoint operator)**.** *Let T be an invertible symmetric operator on H with domain $D(T) \subseteq H$. Then T is self-adjoint if and only if T^{-1} is self-adjoint as well.*

Proof. We will work using the previous graph approach, that is, we want to show the following equivalence

$$\mathcal{G}(T) = V(\mathcal{G}(T))^\perp \;\Leftrightarrow\; \mathcal{G}(T^{-1}) = V(\mathcal{G}(T^{-1}))^\perp.$$

Let us assume the left hand side holds true. Then

$$(y,x) \in \mathcal{G}(T) \Leftrightarrow (y,x) \in V(\mathcal{G}(T))^\perp \;\Leftrightarrow\; \forall\, u \in D(T),\; \langle Tu, y\rangle = \langle u, x\rangle.$$

But by injectivity of T (it is invertible), that is equivalent to saying that

$$\forall\, v \in R(T) = D(T^{-1}),\; \langle v, y\rangle = \langle T^{-1}v, x\rangle \;\Leftrightarrow\; (x,y) \in V(\mathcal{G}(T^{-1}))^\perp.$$

Finally, $(y,x) \in \mathcal{G}(T)$ is equivalent to $(x,y) \in \mathcal{G}(T^{-1})$. Since we derived equivalences, the proof is complete. $\square$

Now something very important to realize is that we mentioned earlier that being everywhere defined was a requirement for our inverse operator. But being everywhere defined and symmetric implies being self-adjoint, since the domain cannot be further extended. In other words, a self-adjoint extension of A is *necessary.*

3.3.4 The space V

In this section, we prove the following theorem.

Theorem 6 (Friedrich Extension)**.** *A positive, densely defined symmetric operator* $(D(A), A)$*, admits a self-adjoint extension.*

To motivate how the proof will go and why the Friedrich extension is suitable, we write this first result.

Proposition 5. *For a self-adjoint operator* $(D(T), T)$ *on* H*, its kernel is the orthogonal complement of its image i.e.* $Ker(T) = R(T)^{\perp}$*.*

Proof. We first show $Ker(T) \subseteq R(T)$.

$$x \in Ker(T) \Rightarrow \langle Tx, u \rangle = 0 \; \forall \, u \in D(T) \Rightarrow \langle x, Tu \rangle = 0 \; \forall \, u \in D(T) \Rightarrow x \in R(T)^{\perp}$$

where we have used the density of $D(T)$ in H. Next, we show $R(T)^{\perp} \subseteq Ker(T)$.

$$y \in R(T)^{\perp} \Rightarrow \langle y, Ax \rangle \; \forall \, x \in D(T) \Rightarrow x \mapsto \langle y, Ax \rangle \text{ is bounded on D(T).}$$

Further, this implies by self-adjointness that

$$y \in D(T^{\star}) = D(T)$$

and $T^{\star} y = Ty = 0$. ☐

Therefore, we want our self-adjoint extension to be injective, as we want its image to be H. In particular, we would like the positivity to remain. We reformulate it here for our operator A:

$$\langle x, Ax \rangle \geq C ||x||^2 \; \forall x \in D(A).$$

The first idea is thus to study the space V defined as

$$V := \{x \in H : \exists \, (x_n) \subseteq D(A), \, ||x_n - x||_H \to 0, \, (x_n) \text{ is cauchy for } || \cdot ||_A\}$$

with

$$|| \cdot ||_V := || \cdot ||_A := \sqrt{\langle \, \cdot \, , A \, \cdot \, \rangle}.$$

That is, we want to close $D(A)$ with respect to this new norm. To show that $|| \cdot ||_A$ will indeed be a norm for V, we need to verify that the limit points x do not depend on the (Cauchy) sequence x_n.

Proposition 6 ([4]). *The norm $||\cdot||_A$ is well defined on V.*

Proof. It is easy to verify symmetry, bilinearity and positive definiteness for the constructed new inner product $\langle\cdot,\,\cdot\rangle_A$. It follows that the norm is well defined on $D(A)$. To show that the extension to these particular limit points are uniquely defined we proceed as follows.

Let $(x_n), (y_n)$ be two Cauchy sequences in $D(A)$ with respect to $||\cdot||_A$, converging to $x, y \in H$ respectively. We get that $z_n := x_n - y_n$ converges to 0 in H and is Cauchy in A. We want to prove that (z_n) is also converging to 0 with respect to $||\cdot||_A$, this will prove that points of V do not depend on the Cauchy sequence and are thus well-defined.

First, note that $||x_n||_A$ is a cauchy sequence in $\mathbb{R}_+$ therefore convergent in $\overline{\mathbb{R}}$.

Ab absurdo, let us assume that $||x_n||_A \to \alpha > 0$.

We observe that

$$\langle x_n, x_m\rangle_A - \langle x_n, x_n\rangle_A = \langle x_n, x_m - x_n\rangle_A \leq ||x_n||_A||x_n - x_m||_A$$

$\forall n, m \in \mathbb{N}$, by bilinearity and Cauchy-Schwarz inequality. Since (x_n) is Cauchy and converges to α with respect to $||\cdot||_A$, we have

$$\forall \epsilon > 0, \exists\, N \in \mathbb{N} \text{ s.t. } \forall\, n, m \geq N, |\langle x_n, x_m\rangle_A - \alpha^2| \leq \epsilon.$$

By taking $\epsilon = \alpha^2/2$, we get

$$|\langle x_n, x_m\rangle_A = |\langle Ax_n, x_m\rangle| \geq \frac{\alpha^2}{2}, \forall n \geq N, \forall m \geq N.$$

Since $x_m \to 0$ in H, $|\langle Ax_n, x_m\rangle|$ goes to zero as $m \to \infty$ leading to a contradiction. □

Hence the space V is well defined. Further, by the Cauchy inequality we also get a natural scalar product on V by

$$\langle u\,,\,v\rangle_V := \lim_{n\to\infty}\langle u_n\,,\,v_n\rangle$$

for $(u_n), (v_n) \subseteq D(A)$ Cauchy sequences for $||\cdot||_V$ tending to u and v respectively.

We further notice that

$$||u||_V \geq C||u||_H\ \forall u \in V$$

as a consequence of the positivity of A. We now have all ingredients to define $A^\star$, our self-adjoint operator. We first define $D(\hat{A})$ by

$$D(\hat{A}) = \{v \in V\ :\ u \mapsto \langle v\,,\,u\rangle_V \text{ is continuous w.r.t. } ||\cdot||_H, \text{ on } V\}.$$

For $v \in D(\hat{A})$, we define $\hat{A}v$ as the unique $w \in H$ such that

$$\langle w \,,\, u \rangle_V = \langle v \,,\, u \rangle_H \; \forall\, u \in V.$$

We indeed have unicity and existence using the Riesz Theorem as well as the density of $V \subseteq H$ (remember that $D(A)$ is contained in V). Let us bound from below this new operator:

$$C||u||_H^2 \leq ||u||_V^2 = |\langle u \,,\, u \rangle_V| = |\langle \hat{A}u \,,\, u \rangle_H| \leq ||\hat{A}u||_H ||u||_H.$$

We therefore have injectivity. Let us now take $h \in H$. We have that $u \mapsto \langle u \,,\, h \rangle_H \; \forall\, u \in V$ is a bounded linear map on V. Therefore by Riesz representation theorem, there exists a $v \in V$ such that $\langle u \,,\, h \rangle_H = \langle u \,,\, v \rangle_V \;\; \forall\, u \in V$. Thus $v \in D(\hat{A})$ and $\hat{A}v = h$. Bijectivity from $D(\hat{A})$ onto H follows. Finally, since the operator is bounded from below, its inverse $\hat{A}^{-1}$ is bounded (from above).

We also have symmetry from the scalar product on V. Its inverse being everywhere defined, it is self-adjoint.

3.3.5 Summary

As a summary of our previous findings, we have seen that the concept of Self-Adjointness is central in order to use the spectral theorem for compact operators. It should be noted that there is in fact a spectral theorem for self-adjoint operators which unfortunately is beyond the scope of this work.

It should also be noted that the previous construction relies heavily on the scalar product. In other words, the choice of scalar product will greatly influence properties of operators. On a side note, the Riesz theorem was central to make sense of points that were necessary to have a well-behaved operator. We rewrite the self-adjoint bijective (on its domain) operator here:

$$\hat{A} : D(\hat{A}) \subseteq V \subseteq H \to H.$$

Finally, there is still a missing piece, namely the compactness embedding. We will show later on that the built space V is indeed compactly embedded in our setting.

3.4 Hilbert Spaces of Functions

We recall that Hilbert spaces are complete normed vector spaces where the norm comes from a scalar product. Now elements from finite or countably infinite dimensional vector space can usually be written as a sequence of (real) numbers $(a_i)_{i \in \mathbb{N}}$ with respect to some basis. The standard scalar product for these spaces is then $\langle a, b\rangle := \sum_{i\in\mathbb{N}} a_i b_i$ whenever it converges and is finite. This natural operation is motivated by its use in geometry and thus physics.

When we consider a space of functions, a basis of functions is not straightforward and writing sums as above is no longer possible. To extend this sum idea in our continuous case, we may use the Lebesgue integral.

Lemma 3 (Delta functions)**.** *Using delta sequences, the Lebesgue integral extends the dot product.*

Proof. For a countable set $S \subset \mathbb{R}$, and continuous functions $f, g \in L^2(\mathbb{R})$, we compute the following limit:

$$I = \lim_{n\to\infty} \int_{-\infty}^{\infty} \sum_{s\in S} \delta_n(x-s) f(x) g(x) dx$$

with

$$\delta_n(x) = \begin{cases} n & \text{if } -\frac{1}{2n} < x < \frac{1}{2n} \\ 0 & \text{otherwise} \end{cases}$$

which is called a delta sequence. Notice how this sequence concentrates the measure around 0. As we will see by computing the limit, this will artificially give a strictly positive measure to the *atom* $\{0\}$ which should not be the case in the Lebesgue measure framework. We shall additionally assume that the sum over the set S is sufficiently well behaved namely that $\sum_{s\in S} f(s)g(s)$ absolutely converges and is finite.

We first change the order of integration and sum. Notice that it can be justified using the dominated convergence theorem (DCT). Indeed, $|\sum \delta \cdot f \cdot g| \leq |nf \cdot g|$ and $\int |nfg| < \infty$ by square integrability of f and g. Therefore,

$$I = \lim_{n\to\infty} \sum_{s\in S} \int_{-\infty}^{\infty} \delta_n(x-s) f(x) g(x) dx.$$

Next, we change the order of limit over n with the sum. This can again be dealt with the DCT only this time we use the counting measure to express the sum as an integral. The hypothesis of the DCT holds since

$$|\int \delta_n \cdot f \cdot g| \leq \int_{s-1/2n}^{s+1/2n} n|f \cdot g| \leq Cn\left(s + \frac{1}{2n} - s + \frac{1}{2n}\right) = C$$

where C is the maximum of $f \cdot g$ over $[s - 1/2n, s + 1/2n]$. Finally we compute the limit of the integral. Notice that by completing the square we can write

$$\int_{-\infty}^{\infty} \delta_n(x-s) f(x) g(x) dx = \int_{s-1/2n}^{s+1/2n} n(f(x)g(x) - f(s)g(s))dx + f(s)g(s).$$

Now, let $\epsilon > 0$. Then by continuity of $f \cdot g$, there exist $N \in \mathbb{N}$ such that

$$\int_{s-1/2n}^{s+1/2n} n(f(x)g(x) - f(s)g(s))dx \leq \int_{s-1/2n}^{s+1/2n} n\,\epsilon\,dx = \epsilon$$

for all $n \geq N$. We thus finally have

$$I = \sum_{s \in S} f(s)g(s).$$

$\square$

Notice that in the above lemma we restricted our attention to a subset of $L^2(\mathbb{R})$, the continuous square integrable functions. This says something about the conditions for which one can jump from a discrete to a continuous setting. Let us present an additional lemma to further motivate our use of the space L^2.

Lemma 4. *L^2 is the unique Hilbert space among L^p spaces.*

Proof. First, recall that for a Lesbegue measurable set X, $L^p(X)$ is defined as the set of all measurable functions f on X for which

$$\int |f|^p d_X$$

is finite. The above quantity can be shown to be well defined as a norm on the $L^p(X)$ space. We then have a norm on a vector space of functions, it is indeed easy to check using Minkowski inequality that linear operations are closed within $L^p(X)$. It can additionally be shown that it is in fact a Banach space, the proof uses measure theory and can be found in any Analysis text book.

Extending the norm to a scalar product, could naively be achieved as follow:

$$\langle f, g \rangle_p := \int |f|^{p/2} |g|^{p/2} d_X. \tag{7}$$

This is indeed the only possible configuration which has the required symmetry property while being a generalization of the above well defined norm.

However, to satisfy the bilinearity property, it should be straightforward that only the choice $p = 2$ is a potential candidate. Finally, we need to take care of the absolute values in equation (7). Notice that when $p = 2$, we do not need them (we actually need to discard them to have bilinearity) as $|f|^2 = f^2$. $\square$

4 The Laplace Operator on $L^2(M)$

We have spent some time to motivate the space $L^2(M)$ above which will be our playground for the Laplace operator. Within this section, our task will be to apply what we have seen on operator theory to H being $L^2(M)$ and A being the Laplace operator.

First, a few notes on the space $L^2(M)$. From the appendix, we have seen how one can integrate smooth functions on a Riemannian manifold. To extend integration to a bigger class of functions, one *completes* the class of compactly contained smooth functions ($C_c^\infty(M)$) using the induced norm from our Hilbert space

$$||f||_{L^2(M)} := \int_M f^2 \mathrm{dvol}_g.$$

This is motivated by the fact that $C_c^\infty(\mathbb{R}^k)$ is indeed densely contained in $L^2(\mathbb{R}^k)$.

4.1 Laplace Operator on a Manifold

Definition (Laplace Operator in $\mathbb{R}^k$)**.** *The Laplace Operator on $f \in C^\infty(\mathbb{R}^k)$ is defined as*

$$\Delta f := -div(\nabla f) = -\sum_{i=1}^{k} \frac{\partial^2}{\partial x^2} f.$$

To extend it to our compact Riemannian Manifold (M, g), we will therefore have to translate the divergence (div : $C^\infty(TM) \to C^\infty(M)$) and the gradient ($C^\infty(M) \to C^\infty(TM)$) operator. We follow the construction given in [6].

First, we denote by α the bundle isomorphism $\alpha : TM \to T^\star M$. Namely, on each fiber $x \in M$ we have $\alpha(v) = v^\star, \forall\, v \in T_x M$ such that $v^\star(w) := \langle v\,,\, w\rangle_{g,x}$. We then define $\nabla : C^\infty(M) \to C^\infty(TM)$ to be the composition

$$C^\infty(M) \xrightarrow{d} C^\infty(T^\star M) \xrightarrow{\alpha_g^{-1}} C^\infty(TM).$$

For the divergence, we may be tempted to define the divergence operator as the adjoint of ∇ with respect to the inner product:

$$\langle -\mathrm{div} X\,,\, f\rangle = \langle X\,,\, \nabla f\rangle, \tag{8}$$

since

$$-\int_{\mathbb{R}^k} \partial_i X^i \cdot f = \int_{\mathbb{R}^k} \partial_i f \cdot x^i$$

for $f \in C_c^\infty(\mathbb{R}^k)$ and $X = X^i \partial_i$ on $\mathbb{R}^k$. Notice that it is important that the functions considered be compactly contained. This is what allows the clean integration by parts above. The right hand side of equation (8) is to be understood as the (global) inner product on TM for the two vector fields X and ∇f, i.e. $\int_M g(X, \nabla f)\, \mathrm{dvol}_g$.

Proposition 7 ([6])**.** *If the div operator indeed exists as the adjoint of the operator* ∇*, then it should be expressed as*

$$div\, X = \frac{1}{\sqrt{det\, g}} \partial_i \left(X^i \sqrt{det\, g} \right)$$

using local coordinates.

Its representation in local coordinates is well defined. So if it is independent of the coordinate system, then div as being the adjoint of ∇ will be well defined.

Proposition 8 ([6])**.** *Given two sets of coordinates* $(x^1, \cdots, x^n)$ *and* $(y^1, \cdots, y^n)$ *on an open set* U *of* M*, we have*

$$\frac{1}{\sqrt{det\, g}} \partial_i \left(X^i \sqrt{det\, g} \right) = \frac{1}{\sqrt{det\, g}} \partial_j \left(Y^j \sqrt{det\, g} \right)$$

with $X = X^i \partial_{x^i}$ *and* $Y = Y^j \partial_{y^j}$.

We now have a well defined Laplace operator on our manifold (M, g). We derive the two following properties:

Symmetry. By the adjoint property of the div operator we have $\langle \Delta f\, ,\, g \rangle = \langle -\mathrm{div}\,(\nabla f)\, ,\, g \rangle = \langle \nabla f\, ,\, \nabla g \rangle = \langle f\, ,\, \Delta g \rangle$ for all $f, g \in C_c^\infty(M)$.

Positivity. It comes easily from the symmetry: $\langle \Delta f\, ,\, f \rangle = \langle \nabla f\, ,\, \nabla f \rangle \geq 0$ for all $f \in C_c^\infty(M)$.

Thus, we may now say that

$$\Delta : C_c^\infty(M) \subseteq L^2(M) \to L^2(M)$$

has the required properties to use the theory from the previous section. In particular, the space V will be the completion of $C_c^\infty(M)$ with respect to the norm

$$||f||_V^2 = \langle \Delta f\, ,\, f \rangle = \langle \nabla f\, ,\, \nabla f \rangle = ||\nabla f||_{L^2(M)}^2.$$

Notice that the norm that defines V controls the behavior of the function's growth which should intuitively feel stronger for functions defined on a bounded set.

4.2 Sobolev spaces

Let us first work in $\Omega \subseteq \mathbb{R}^k$ a bounded and open set. For $s \in \{0, 1, 2, \cdots\}$ we denote by $H^s(\Omega)$ the completion of $C_c^\infty(\Omega)$ with respect to the norm

$$||f||_s := (\sum_{|\alpha| \leq s} ||D^\alpha f||_2^2)^{\frac{1}{2}}) \tag{9}$$

with $\alpha = (\alpha_1, \cdots, \alpha_k), \alpha_j \in \mathbb{Z}$ a multi-index, such that $|\alpha| := \sum_j \alpha_j$ and

$$D^\alpha := (-i)^{|\alpha|} \frac{\partial^{\alpha_1 + \cdots + \alpha_k}}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_k}}.$$

Notice that with this new notation, $L^2(\Omega)$ can be written as $H^0(\Omega)$. To avoid confusion with the norms, we will denote the L^2 norm by $|| \cdot ||_{L^2}$. Further we have included an optional imaginary factor which should only be present when considering complex valued functions (such as in Fourier analysis).

The following results will rely heavily on the Fourier transform. It indeed allows us to work in an alternative space with much more ease. It is however not necessary and we refer the reader to [8] for the following results without using the Fourier transform. We assume results about Fourier analysis to be known, and recall that for $u \in L^2(\Omega)$ we have

$$\hat{u}(\xi) = \frac{1}{(2\pi)^k} \int_{\mathbb{R}^k} e^{-ix\xi} u(x) dx.$$

Lemma 5 (Equivalence of Norms [6]). *For $f \in C_c^\infty(\Omega)$, we have*

$$||f||_s \approx \left(\int_{\mathbb{R}^k} |\hat{f}(\xi)|^2 (1 + |\xi|^2)^s d\xi \right)^{\frac{1}{2}}.$$

The approximation symbol between two norms $|| \cdot ||_V$ and $|| \cdot ||_W$ means that there exists C_1, C_2 such that $C_1 || \cdot ||_V \leq || \cdot ||_W \leq C_2 || \cdot ||_V$.

sketch of proof.
The first step is to develop $(|| \cdot ||_s)^2$ and using the isometry property of the Fourier transform. The second step uses the following fact:

$$C_1 (1 + |\xi|^2)^s \leq \left(\int_{|\alpha| \leq s} |\xi|^\alpha \right)^2 \leq C_2 (1 + |\xi|^2)^s$$

for some $C_1, C_2 \in \mathbb{R}_+$. □

Lemma 6 (Continuous embedding of Sobolev spaces [6]). *For $s > t \in \mathbb{N}$, we have*

$$H^s \hookrightarrow H^t$$

or in other terms, $C|| \cdot ||^s \geq || \cdot ||^t$ for a positive constant C.

Proof. Using equation (9), one observes that positive terms are being added as s grows. Hence we have

$$||f||_t \leq ||f||_s \, , \forall f \in H^s.$$

This implies the continuous embedding. $\square$

Theorem 7 (Poisson equation). *For $\mu > 0$ and $f \in L^2(\Omega)$, there exists a unique $u \in H^2(\Omega)$ such that*

$$\Delta_\mu u = f \tag{10}$$

where $\Delta_\mu := \Delta + \mu I_d$. Furthermore, there exist positive constants C_1, C_2 such that

$$C_1 ||f||_{L^2} \leq ||u||_2 \leq C_2 ||f||_{L^2}.$$

Proof. We recall here that the extended operator

$$\hat{\Delta}_\mu : D(\hat{\Delta}_\mu) \subseteq V \subseteq L^2(\Omega) \to L^2(\Omega)$$

is bijective onto $L^2(\Omega)$. Hence, since $f \in L^2(\Omega)$, the solution exists (surjectivity) and is unique (injectivity) and belongs to V. This solution is called weak, because we do not know yet if equation (10) holds true almost everywhere (a.e.). Indeed, we have only found a solution for our extended operator.

Our next goal is thus to improve the regularity of our solution u. Let us bound the H^2 norm of u.

$$\begin{aligned} ||u||_2 \approx (||\xi^2 \hat{u}||_{L^2}^2 + ||\hat{u}||_{L^2}^2) &= ||\frac{\partial^2}{\partial x_j^2} u||_{L^2}^2 + ||u||_{L^2}^2 \\ &\leq C_\mu (||\sum_{i=1}^{k} \frac{\partial^2}{\partial x_i^2} u||_{L^2}^2 + ||\mu u||_{L^2}^2) \qquad (11) \\ &= C_\mu ||(\Delta + \mu I_d) u||_{L^2}^2 \\ &= C_\mu ||f||_{L^2}. \end{aligned}$$

Thus, $u \in H^2(\Omega)$.

Finally, to show the last inequality, we take a second look at the inequation (11). Let us name the largest contributor over the k second derivatives by $j \in \{1, \cdots, k\}$. we then have that

$$C_\mu (||\sum_{i=1}^{k} \frac{\partial^2}{\partial x_i^2} u||_{L^2}^2 + ||\mu u||_{L^2}^2) \leq \hat{C} k (||\frac{\partial^2}{\partial x_j^2} u||_{L^2}^2 + ||u||_{L^2}^2)$$

for some $\hat{C} \in \mathbb{R}_+$. This proves that we have $||u||_2 \approx ||f||_{L^2}$. $\square$

Theorem 8 (Rellich-Kondarachov Compactness Theorem [6]). *If $s > t \in \mathbb{N}$, then the inclusion $H^s(\Omega) \to H^t(\Omega)$ is compact.*

Proof. Let us take a sequence $(f_n) \subseteq H^s$ such that $||f_n||_s \leq 1$ for all $n \in \mathbb{N}$. To prove the compactness embedding, we want to exhibit a subsequence that is Cauchy within the space H^t (i.e. with respect to the norm $|| \cdot ||_t$). Notice that for a fixed $r > 0$

$$\begin{aligned} ||f_n - f_m||_t^2 &\approx \int_{\mathbb{R}^k} |\hat{f}_n(\xi) - \hat{f}_m(\xi)|^2 (1 + |\xi|^2)^t d\xi \\ &= \int_{|\xi| \leq r} |\hat{f}_n(\xi) - \hat{f}_m(\xi)|^2 (1 + |\xi|^2)^t d\xi \\ &\quad + \int_{|\xi| > r} |\hat{f}_n(\xi) - \hat{f}_m(\xi)|^2 (1 + |\xi|^2)^t d\xi. \end{aligned}$$

We may now use the fact that $s > t$ and $|\xi| > r$ in the second term to write

$$\begin{aligned} \int_{|\xi| > r} |\hat{f}_n(\xi) - \hat{f}_m(\xi)|^2 (1 + |\xi|^2)^t d\xi &\leq \frac{1}{(1 + r^2)^{s-t}} \int_{|\xi| > r} |\hat{f}_n(\xi) - \hat{f}_m(\xi)|^2 (1 + |\xi|^2)^s d\xi \\ &\leq \frac{1}{(1 + r^2)^{s-t}} ||f_n - f_m||_s^2 \\ &\leq \frac{4}{(1 + r^2)^{s-t}}. \end{aligned}$$

We can then send it to zero using r.

Now notice that if we have uniform convergence for $(\hat{f}_n)$ on compact sets (such as on $\{\xi \in \mathbb{R}^k : |\xi| \leq r\}$), the second term can also be bounded. Then, for every $\epsilon > 0$, we can choose $r > 0$ to have the first term bounded by $\epsilon/2$ and then $n, m > N \in \mathbb{N}$ to bound the second term by another $\epsilon/2$.

To show uniform convergence, the Arzela-Ascoli theorem comes handy as it will give existence of such a subsequence under some assumptions. To verify these assumptions, it is sufficient to bound $|\frac{\partial}{\partial \xi_i} \hat{f}_n|$ for $i \in \{1, \cdots, k\}$ to get equicontinuity, and to bound $|\hat{f}_n|$ to get uniform boundedness (which are easy to derive). Those bounds can depend continuously on ξ, since it will be enough to bound the expressions on the compact sets $\{\xi \in \mathbb{R}^k : |\xi| \leq r\}$. $\square$

4.3 From Ω to M

We have worked with an open bounded set in the previous section to understand Sobolev spaces. We now transfer these spaces as well as their results to M a compact riemannian connected k-manifold. We additionally assume that M has no boundary.

Again, we follow [6] with respect to the notation and definitions.

Definition (Sobolev Spaces on Manifolds [6])**.** *Let us take $\{(\varphi_i, U_i)\}_{i \in I}$ a locally finite coordinate cover of M such that $\varphi_i : U_i \subseteq \mathbb{R}^k \to M$ with $\overline{U_i}$ compact for all*

$i \in I$. We additionally take $\{\rho_i\}_{i\in I}$ a partition of unity subordinate to U_i. We set $H^s(M)$ to be the completion of $C_c^\infty(M)$ with respect to

$$||f||_s := (\sum_{i\in I} ||(\rho_i f)\circ\varphi_i||_s^2)^{1/2}.$$

Notice that $((\rho_i f)\circ\varphi_i)$ is indeed a smooth function with compact support. To show that the space $H^s(M)$ does not depend on the coordinate system, it is sufficient to show that two norms using different coordinate systems are equivalents. Thus, let us consider two triplets $\{(U_i, \varphi_i, \rho_i)\}_{i\in I}$, $\{(V_j, \psi_j, \mu_j)\}_{j\in J}$ respectively with the open locally finite cover, associated (inverted) chart and the subordinated partition of unity.

The first thing to note is that a locally finite cover on a compact manifold implies that the cover has a finite number of elements. In other words $|I|, |J| < \infty$.

The second thing to note is that derivatives and partitions of unity commute. Indeed, for a multi index α such that $|\alpha| = 1$, we have

$$\frac{\partial}{\partial x^\alpha}\left(\sum_i \rho_i f\right) = \frac{\partial}{\partial x^\alpha} f = \sum_i \rho_i \left(\frac{\partial}{\partial x^\alpha} f\right)$$

for f having a compact domain and $\{\rho_i\}$ a smooth partition of unity subordinated to some open cover of this compact domain.

For a fixed $i \in I$ we define

$$W_j := \psi_j(V_j) \cap \varphi_i(U_i) \subseteq M$$

for all those $j \in J$ such that $W_j \neq \emptyset$ (a set $S_i \subseteq J$). Notice that $\varphi_i(U_i) \subseteq \bigcup_{j\in S_i} W_j$ since $\{\psi_j(V_j)\}_{j\in J}$ is an open cover of M.

Next, we define the diffeomorphism

$$\begin{aligned}\tau_j : \psi_j^{-1}(W_j) &\to \varphi_i^{-1}(W_j)\\ x &\mapsto (\varphi_i^{-1}\circ\psi_j)(x).\end{aligned}$$

We then have

$$||(\rho_i f)\circ\varphi_i||_s^2 = \sum_{j\in S_i} ||(\rho_i\mu_j f)\circ\varphi_i||_s^2 = \sum_{j\in S_i} ||(\rho_i\mu_j f)\circ\psi_j \det(\tau_j)||_s^2$$

where we have multiplied by the second partition of unity $\{\mu_j\}_{j\in J}$. Further, since $J, \det(\tau_j) < \infty$ there exists $C_i > 0$ and $j \in J$ such that

$$\sum_{j\in S_i} ||(\rho_i\mu_j f)\circ\psi_j \det(\tau_j)||_s^2 \le C_i||(\rho_i\mu_j f)\circ\psi_j||_s^2 \le C_i||(\mu_j f)\circ\psi_j||_s^2.$$

By going through a similar process for $||(\mu_j f) \circ \psi_j||_s^2$, we can find $C_1, C_2 > 0$ such that

$$C_1 \left(\sum_{i \in I} ||(\rho_i f) \circ \varphi_i||_s^2 \right) \leq \sum_{j \in J} ||(\mu_j f) \circ \psi_j||_s^2 \leq C_2 \left(\sum_{i \in I} ||(\rho_i f) \circ \varphi_i||_s^2 \right).$$

We now have a coordinate independent definition for Sobolev spaces on compact manifolds. The final result we prove here will be to adjust the compact embedding theorem for Ω on our manifold M.

Theorem 9 (Compactness Theorem for Compact Manifolds [6])**.** *If $s > t \in \mathbb{Z}$, then the inclusion $H^s(M) \to H^t(M)$ is compact.*

Proof. Let us take a sequence $\{f_j\} \subseteq H^s(M)$ with $||f_j||_s$ bounded. Notice that in the ith ($i \in I$) coordinate chart U_i, $\{\rho_i f_j \varphi\}$ satisfies the hypothesis of the Rellich-Kondarachov theorem (8). It therefore has a convergent subsequence in $H_t(U_i)$. Since I is finite, we may build a subsequence - that we rename $\{\tilde{f}_j\}$ - which is convergent in $H_t(U_i)$ for all $i \in I$. Finally, for $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n, m > N$ and all $i \in I$

$$||(\rho_i(\tilde{f}_n - \tilde{f}_m)) \circ \varphi_i||_t^2 < \epsilon/I,$$

leading to

$$||\tilde{f}_n - \tilde{f}_m||_t^2 = \sum_{i \in I} ||(\rho_i(\tilde{f}_n - \tilde{f}_m)) \circ \varphi_i||_t^2 < \epsilon.$$

□

Theorem 10 (Elliptic Proof of the Hodge Theorem for Functions)**.** *Let*

$$\Delta : C_c^\infty(M) \subseteq L^2(M) \to L^2(M)$$

be our Laplace operator. Then there exists an orthonormal basis of $L^2(M)$ consisting of eigenfunctions for the Laplace operator.

Proof. We paste our previous findings.

First, the Laplace operator can be extended to be surjective onto $L^2(M)$, using the Friedrich extension. Next, from the compactness embedding theorem, its domain is compactly embedded in $L^2(M) = H_0$. This makes the extended inverse a self-adjoint compact operator. We may now use the spectral theorem for compact operators to conclude.

Notice that the injectivity of the operator is resolved by isolating the kernel, which corresponds to the eigenvalue 0. Further, this eigenspace is easy to determine. If $\Delta f = 0$, then $0 = \langle \Delta f, f \rangle = \langle \nabla f, \nabla f \rangle$. Since we consider a connected manifold, f must be a constant function making the kernel a one dimensional space. □

We have not mentioned the Sobolev embedding theorem, yet together with theorem (7) this tells us that the eigenfunctions will be smooth. Indeed, from (7), we can iterate on the regularity of the solution for the Poisson equation. This will mean that an eigenfunction ϕ_i will belong to $\cap_{k\in\mathbb{N}} H_k(M)$. Using the Sobolev embedding theorem, we then have that $\phi_i \in C^\infty(M)$.

5 Conclusion

The core of the work has consisted of proving the spectral decomposition of the Laplace operator on a Compact Riemannian Manifold. We started by using a general approach with functional analysis. This allowed to get almost all the hypothesis required to apply the spectral theorem for compact operators. To get the last missing piece - the compactness embedding - we went through Sobolev spaces to then use the Rellich-Kondarachov theorem.

A
Differential Geometry

We try to give here sufficient background and motivation for the following result:

Theorem 11 (Stokes Theorem)**.** *Let M be an oriented smooth n-manifold with boundary, and let ω be a compactly supported smooth $(n-1)$-form on M. Then*

$$\int_M d\omega = \int_{\partial M} \omega.$$

A.1 "M an oriented smooth n-manifold with boundary"

A.1.1 Smooth Manifolds

One motivation for (smooth) manifolds is to extend the differential geometry of surfaces. Let us first notice that differentiability is a local property where "local" can be defined using a topology. Further, we would like to use the findings of real analysis but in our more general setting. One way of doing this is through local *diffeomorphisms*. Indeed, if there exists $k \in \mathbb{N}$ such that opens of our set are diffeomorphic to an open set of $\mathbb{R}^k$ we can then transfer results through this morphism.

Definition (Topological Manifold [9])**.** *Let M be a topological space. M is a topological manifold of dimension n if it has the following properties.*

- *M is a Hausdorff space.*
- *M is second countable.*
- *Each point of M has a neighborhood that is homeomorphic to an open subset of $\mathbb{R}^n$.*

Definition (Smooth Charts [9])**.** *The homeomorphisms mentioned in the definition of a topological manifold are called charts. They come as a pair (U, φ) such that $\varphi : U \subseteq M \to \hat{U} \subseteq \mathbb{R}^n$ is a homeomorphism for some open $\hat{U} \subseteq \mathbb{R}^n$. From the definition, they cover the whole space M. Now we say they are smoothly compatible if for any two pairs $(U, \varphi), (V, \psi)$ either $U \cap V = \emptyset$ or $U \cap V \neq \emptyset$ and*

$$\psi \circ \varphi^{-1} : \varphi(U \cap V) \subseteq \mathbb{R}^n \to \psi(U \cap V) \subseteq \mathbb{R}^n$$

is a diffeomorphism. The composition above is called a transition map from φ to ψ. The collection of such charts is called a smooth atlas.

Definition (Smooth Manifold)**.** *An atlas $\mathcal{A}$ on M is maximal if it is not properly contained in any larger smooth atlas. In other words, any charts that is smoothly compatible with every chart in the atlas is already in it. Such $\mathcal{A}$ is also called a smooth structure. Now a smooth Manifold is a pair $(M, \mathcal{A})$ with M a topological manifold and $\mathcal{A}$ a smooth structure.*

As there is a topology associated to M, so is the compactness property from standard topology. Now for $f : M \to \mathbb{R}$, we say that f is smooth at $p \in M$ if for any open U containing p and the associated chart φ_U,

$$f \circ \varphi_U^{-1} : \varphi_U(U) \subseteq \mathbb{R}^k \to \mathbb{R}$$

is smooth in the standard calculus sense.

A.1.2 Boundaries

Now to make sense of boundaries, we use the space

$$\mathbb{H}^n := \{(x^1, \cdots, x^n) \in \mathbb{R}^n : x^n \geq 0\}$$

as well as *boundary charts*, charts φ belonging to the atlas such that $\varphi(U)$ is an open subset of $\mathbb{H}^n$ (relative to $\mathbb{R}^n$) with $\varphi(U) \cap \partial\mathbb{H}^n \neq 0$. We can then define boundary points of M as being the points $p \in U$ with $\varphi(p) \in \partial\mathbb{H}^n$. Notice that choosing the last component in the definition of $\mathbb{H}^n$ is arbitrary, as an argument permutation in φ does not invalidates its belonging to the atlas.

A.1.3 Orientation

We will deal with orientation while talking about integration.

A.2 "ω a compactly supported smooth $(n-1)$-form on M"

A.2.1 Vector Fields

Let us first talk about vector fields and the space in which they live in. Let us take $p \in \mathbb{R}^n$ and $f \in C^\infty(\mathbb{R}^n)$ the space of infinitely differentiable functions on $\mathbb{R}^n$. Then a vector field $X : \mathbb{R}^n \to \mathbb{R}^n$ acts on $C^\infty(\mathbb{R}^n)$ by returning the derivative of f in the direction $X(x)$ at x:

$$\begin{aligned} X(x) : C^\infty(\mathbb{R}^n) &\to \mathbb{R} \\ f &\mapsto \langle X(x), df(x) \rangle =: (X \cdot f)(x) \end{aligned}$$

using the standard scalar product.

Notice that we have loosely used the same term X for the linear functional as well as for the collection of vectors. Let us link the two concepts by developping the scalar product:

$$\langle X(x),\, df(x)\rangle = \sum_{i=1}^{n} X^i(x)\frac{\partial}{\partial x^i}f(x) \tag{12}$$

with $(X^i)_{i=1,\cdots,n}(x)$ the components of the vector $X(x)$. Further, notice that the functionals $(\frac{\partial}{\partial x^i})_{i=1,\cdots,n}(x)$ span the tangent space of $\mathbb{R}^n$ at x where the derivatives of functions at x live.

For every vector fields, we have a linear functional $C^\infty(\mathbb{R}^n) \to \mathbb{R}$. It turns out we can characterize this mapping using the standard *product rule* of calculus.

Proposition 9 ([9])**.** *A linear functional $X : C^\infty(\mathbb{R}^n) \to \mathbb{R}$ satisfying the product rule $X \cdot (fg) = (X \cdot f)g + (X \cdot g)f$ is a vector field on $\mathbb{R}^n$.*

Note that the product rule equality has to be tested with $x \in \mathbb{R}^n$. Now the idea is to generalize vector fields on $\mathbb{R}^n$ over a manifold M using the above proposition. That is, a vector field on M is any linear functional $X : C^\infty(M) \to \mathbb{R}$ satisfying the product rule. Notice that this definition does not involve any charts and is thus *coordinate independent*.

We call the space in which vector fields on a manifold M live, a *tangent bundle* $TM := \cup_{p\in M} T_pM$ the bundle of all tangent spaces of M. The tangent space at point p here means the space of all vector fields at point p as defined using functionals.

Next, we express X using a coordinate system. First, we define $\tilde{X}$ on $\mathbb{R}^n$ as

$$\begin{aligned}\tilde{X}(x) : C^\infty(\mathbb{R}^n) &\to \mathbb{R}\\ f &\mapsto (X \cdot f \circ \varphi)(p) = (X \cdot \tilde{f})(p)\end{aligned}$$

with $\varphi^{-1}(x) = p$ using the suitable chart & open couple around p. It is easy to check that $\tilde{X}$ satisfies the product rule, it is thus a vector field on $\mathbb{R}^n$. As such, for a given $x \in \mathbb{R}^n$, we are able to express X in coordinates using the formulation of equation (12) with $\tilde{f}$.

A.2.2 Covariant Tensors

We have seen how vector fields act on smooth functions through the differentiation in a given direction. Conversely, smooth functions act on vector fields using the same operation.

$$\begin{aligned}f(p) : T_pM &\to \mathbb{R}\\ X(p) &\mapsto \langle X(p),\, df(p)\rangle = (X \cdot f)(p).\end{aligned}$$

In general, a covariant one-tensor field is any linear map

$$\begin{aligned}\alpha(p) : T_pM &\to \mathbb{R}\\ X(p) &\mapsto \alpha(p)(X(p))\end{aligned}$$

For all $p \in M$.

A nice way of apprehending covariant tensor fields on a manifold is to build them with coordinates. Notice that we hinted at $(\frac{\partial}{\partial x^i})_{i=1,\cdots,n}(p)$ being a coordinate system for vector fields on a manifold, let us similarly use the coordinates $(x^i)_{i=1,\cdots,n}(p)$ as smooth functions. Let us define $dx^i(p)$ as a mapping from T_pM to $\mathbb{R}$.

$$\langle X(p), dx(p)\rangle = \sum_{i=1}^{n} X^i(p)\frac{\partial}{\partial x^i}x(p) =: \sum_{i=1}^{n} X^i(p)dx^i(p).$$

dx^i is an example of tensor, a covariant one-tensor field to be precise since it will take as argument one vector field $X(p)$ and return its component $X^i(p)$. By convention, (smooth) covariant zero-tensor fields are smooth functions on M and more generally

$$\begin{aligned}\beta(p) : T_pM \times \cdots \times T_pM &\to \mathbb{R}\\ (X_1(p), \cdots, X_n(p)) &\mapsto \beta(p)(X_1(p), \cdots, X_n(p))\end{aligned}$$

is a covariant n-tensor field on M.

We have introduced covariant one-tensor fields by way of smooth functions acting on vector fields. Notice that we could introduce some covariant two-tensor fields using the same idea, namely

$$\begin{aligned}f(p) : T_pM \times T_pM &\to \mathbb{R}\\ (X_1(p), X_2(p)) &\mapsto X_1(p)^T\, d^2f(p)\, X_2(p) = \sum_{i,j=1}^{n} X_1^i X_2^j \frac{\partial^2}{\partial x^i \partial x^j} f(p).\end{aligned}$$

This forces the covariant two-tensor to be *symmetric*. Indeed, permuting arguments has no effect since the matrix $d^2f(p)$ is symmetric by the smoothness of f. We next see an opposite class of tensors, namely the alternating ones which on the other hand will switch signs if arguments are switched. In what follows we use the word tensor to describe covariant tensors. By doing so, we intentionally put aside *contravariant* as well as mixed types tensors [9].

A.2.3 Motivation for Alternating Tensors

Say we have a smooth function f on our manifold M assumed now to be compact, which we would like to integrate: $I = \int_M f$. When $M = [a_1, b_1] \times \cdots \times [a_n, b_n], a, b \in \mathbb{R}^n$ we know from standard calculus that

$$I = \int_{[a_1,b_1]\times\cdots\times[a_n,b_n]} f = \int_{[a_1,b_1]\times\cdots\times[a_n,b_n]} f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n$$

using the obvious coordinate system. In the Riemannian framework, integrals need *volume elements* together with the limit of a sum. For example $[a_1, b_1]$ gets subdivided in a collection of small segments, $[a_1, b_1] \times [a_2, b_2]$ by small squares, $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$ by cubes and so on and so forth.

The decisive observation that will allow us to generalize is that the volume of these elements are given by an alternating tensor, the *determinant*. Indeed, by rewriting the above n-cubes as spanned by n vectors, stacking the vectors in a matrix and computing its determinant, this will give their volumes.

Hidden in the expression for I, is the determinant of the standard coordinate representation of $\mathbb{R}^n$ which is equal to 1. Recall that when a change of coordinates occur, a Jacobian appears which gives the value of the determinant with respect to the coordinate transformation (a linear map).

In other terms, another way of looking at the standard integration theory is to say that we integrate tensor fields, where each element of the sum is the value of the tensor evaluated in infinitesimal volume elements spanned by a set of vectors.

Let us study properties of the determinant to understand a possible generalization.

Definition (Determinant [9]). *For a finite n-dimensional vector space V together with an orthonormal basis B, a linear map $T : V \to V$ written as a matrix $(T)_{i,j}$ with respect to the basis B has determinant*

$$det_B(T) = \sum_{\sigma \in S_n} sgn(\sigma) \prod_{i=1}^{n} T_{i,\sigma(i)}.$$

By looking at the matrix $(T)_{i,j}$ as an ordered collection of vectors from V, we can further write det_B as a mapping

$$det_B : V \times \cdots \times V \to \mathbb{R}.$$

It can further be shown that the following properties for the mapping det_B hold:

Multilinear For all $\lambda, \beta \in \mathbb{R}$ and all $(v_i)_{i=1,\cdots,n}, w \subseteq V$,

$$\begin{aligned} det_B(v_1, \cdots, \lambda v_i + \beta w, \cdots, v_n) &= \lambda det_B(v_1, \cdots, v_i, \cdots, v_n) \\ &\quad + \beta det_B(v_1, \cdots, w, \cdots, v_n). \end{aligned}$$

Alternating For all $(v_i)_{i=1,\cdots,n} \in V$,

$$det_B(v_1, \cdots, v_i, \cdots, v_j, \cdots, v_n) = -det_B(v_1, \cdots, v_j, \cdots, v_i, \cdots, v_n).$$

$det_B(B) = 1$ That is, writing the orthonormal basis B as $(e_i)_{i=1,\cdots,n} \subseteq V$, we have

$$det_B(e_1, \cdots, e_n) = 1.$$

To get a feel about the alternating property, notice that it is equivalent to say that if two vectors v_i, v_j are dependent, then $det_B(v_1, \cdots, v_i, \cdots, v_j, \cdots, v_n)$ is zero. This is coherent with the idea of volumes and their sizes. Indeed, to describe a volume with non zero size, the vectors spanning the volume have to be independent otherwise the volume gets "flattened".

Now what will help us to generalize is the observation that it is possible to characterize the determinant using the above properties. Namely, any covariant alternating n-linear map ω on V satisfying $\omega(e_1, \cdots, e_n) = 1$ is a determinant with respect to the basis B.

We talked about the Jacobian before, the next proposition states it rigorously.

Proposition 10 (Jacobian [9])**.** *Let us take an n-dimensional vector space V and an n-multilinear and alternating tensor ω. If $T : V \to V$ is any linear map and $v_1, \cdots, v_n$ arbitrary vectors in V, then*

$$\omega(Tv_1, \cdots, Tv_n) = det(T)\, \omega(v_1, \cdots, v_n).$$

Thus far, for an n dimensional manifold, we have seen vector fields acting on smooth functions, smooth functions acting on vector fields, we have seen an example of symmetric 2-tensor field (the Hessian matrix), and finally the determinant as an example of alternating n-tensor. Remember that the determinant is well defined on tangent spaces of our manifold but is not yet a tensor field. Before continuing with operations on tensors, let us introduce the notion of forms to ease the reading.

Definition. *A k-form is a smooth alternating covariant k-tensor field.*

[TO DO: smoothness of forms]

To make sense of the Stokes theorem, we must introduce the *exterior derivative* of a k-form ω.

[Wedge product]

[Exterior derivative]

[It became too time consuming for too little knowledge return to keep on going.]

A.3 Integration on a Manifold

Now let us give a few comments on the *Stokes Theorem* that we recall here:

Theorem 12 (Stokes Theorem [9]). *Let M be an oriented smooth n-manifold with boundary, and let ω be a compactly supported smooth $(n-1)$-form on M. Then*

$$\int_M d\omega = \int_{\partial M} \omega.$$

Notice how the result generalizes the fundamental theorem of calculus. Indeed, notice that when $M = [a, b]$, then ω is a smooth function (zero-form) which we denote by F and by f its derivative. Then,

$$\int_{[a,b]} f = F(b) - F(a)$$

which takes the points a, b as the boundaries of $[a, b]$. It also generalizes the divergence theorem, green theorem and Kelvin-Stokes theorem [9].

References

[1] Stéphane Lafon, Diffusion Maps and Geometric Harmonics, PhD dissertation, Yale University, 2004.

[2] Trevor Hastie, Principal Curves and Surfaces, PhD dissertation, Stanford University, 1984.

[3] A. N. Gorban, B. Kégl, D. C. Wunsch, A. Zinovyev, Principal Manifolds for Data Visualization and Dimension Reduction, Springer, 2008, pp 97.

[4] Bernard Helffer, Spectral theory and applications. An elementary introductory course. Bucarest, 2010.

[5] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation, 15(6) 1373 1396, June 2003.

[6] S. Rosenberg, The Laplacian on a Riemannian Manifold, Cambridge Univ. Press, 1997.

[7] B. Buffoni, Analyse Fonctionnelle, Lecture Notes, EPFL, 2012.

[8] L. C. Evans, Partial Differential Equations, American Mathematical Society, 2010.

[9] John M. Lee, Introduction to Smooth Manifolds, Springer, New York, 2013.