POLITECNICO DI MILANO

SCHOOL OF INDUSTRIAL AND INFORMATION ENGINEERING,
IN COLLABORATION WITH THE NATIONAL RESEARCH INSTITUTE INRIA
Master's degree in Mathematical Engineering

WHITE MATTER TRACTOGRAPHY

Application of shape analysis and functional data analysis tools
on fiber bundles analysis

Candidate:
Lorenzo Rota
Personal ID 838167

Thesis supervisor:
Simone Vantini

Research supervisors:
Olivier Commowick
Christian Barillot
Aymeric Stamm

Academic Year 2016-2017

Acknowledgements

This work represents the closure of an important part of my life, full of experiences and challenges, but above all characterized by a positive personal growth. The decision to do my thesis abroad was motivated by the desire to put myself on the line. I am grateful to the VISAGES team, especially to Olivier and Christian who gave me the right suggestions along my internship. I would like to thank Simone for giving me this opportunity and for his help as my supervisor, together with Aymeric.
During my stay in France i also experienced rough time, i owe my success mostly to my family and to my girlfriend, Laura, who have always been close to me along all my stay abroad. I will be always thankful to my parents, Ugo and Lorella, without their help and their guidance i would have never reached my objectives. They encouraged me in every important decision along my route.
Lastly, i should like to thank also my friends, those who lived the university career with me and those who have been always an important moral support.

Contents

Introduction xiii

Introduzione xvii

1 Whitte Matter Tractography 1
- 1.1 Diffusion Magnetic Resonance . . . 1
  - 1.1.1 Principles of Diffusion . . . 1
  - 1.1.2 Underpinnings of Diffusion MR . . . 2
  - 1.1.3 Estimation of the Apparent Diffusion Coefficient . . . 4
- 1.2 Diffusion Tensor Model . . . 5
  - 1.2.1 Model Building . . . 5
  - 1.2.2 Scalar Measures From the Diffusion Tensor . . . 5
  - 1.2.3 Advantages and Limitations of DTI . . . 7
- 1.3 Tractography . . . 7
  - 1.3.1 Deterministic Tractography . . . 8

2 Multiple Sclerosis Research Study 11
- 2.1 Application of DTI in MS . . . 11
- 2.2 Preprocessing of the data . . . 12
- 2.3 Research Goals . . . 14

3 Shape Analysis 17
- 3.1 Differentiable Manifolds . . . 18
  - 3.1.1 Definition . . . 18
  - 3.1.2 Tangent space . . . 21
  - 3.1.3 Extension to the Infinite-dimensional Case . . . 23
- 3.2 Riemannian Framework . . . 24
  - 3.2.1 Riemannian Metric and Geodesics . . . 24
  - 3.2.2 Group Actions and Quotient Space . . . 27
- 3.3 Representation of Curves . . . 28
  - 3.3.1 Definition and Motivation . . . 28
  - 3.3.2 Preshape and Shape spaces under SRVF . . . 30
  - 3.3.3 Geodesics in the representative space . . . 31
  - 3.3.4 Joint Analysis with other Features . . . 33

4 Applications in Statistical Modeling of Fibers 37
- 4.1 Summary Statistics . . . 37
- 4.2 Hierarchical Clustering . . . 39
- 4.3 Registration . . . 40

4.4 Summary 43

5 A New Perspective with Functional Data Analysis 45
5.1 Goals and Proposals 45
5.2 Preparing the Data 47
5.3 Interval-Wise Testing Procedure 48
5.3.1 Detection of a Clinically Significant Threshold 49
5.4 Functional Linear Discriminant Analysis 50
5.4.1 Applications 53

6 Conclusions and Perspectives 57

Bibliography 59

List of Figures

1.1 1D Gaussian diffusion . . . 2
1.2 PGSE sequence . . . 3
1.3 PGSE measurements . . . 4
1.4 Configurations of the tensor model . . . 6
1.5 Example of scalar measures from diffusion . . . 7
1.6 FACT . . . 8
1.7 Example of brain tractography . . . 9

2.1 Lesions on a T_2 image . . . 12
2.2 Anatomy of the CST . . . 13
2.3 Tracts extracted with WMQL . . . 14
2.4 Example of brain parcellation . . . 15

3.1 Example of charts on a manifold . . . 20
3.2 Example of a smooth map on a manifold . . . 21
3.3 Example of tangent space . . . 24
3.4 Geodesics on the earth . . . 26
3.5 Exponential map . . . 27
3.6 Geodesic optimization on shape space . . . 32
3.7 Geodesic optimization with different features . . . 34

4.1 Example of Karcher means . . . 38
4.2 Example of Wrapped Gaussian resampling . . . 39
4.3 Principal paths for the first eigenmode . . . 40
4.4 Variation along the first two eigenmodes . . . 40
4.5 Example of clustering . . . 41
4.6 Registration in the augmented space . . . 42
4.7 Groupwise registration . . . 44

5.1 Example of a FA profile along a CST . . . 46
5.2 Lesion on a CST . . . 46
5.3 Population Mean . . . 47
5.4 ROC curve . . . 51
5.5 IWT results . . . 52
5.6 GCV curve . . . 53
5.7 FA functions fitted with FLDA . . . 54
5.8 Centroids . . . 54
5.9 Discriminant Function . . . 55
5.10 Classification performance . . . 56

List of Tables

3.1 Representative spaces 36

5.1 Confusion matrix 50
5.2 Missclassification Table 55

Abstract

In a context of a quantitative analysis of white matter fibers, there is a growing need to develop statistical methods able to study diffusion properties along major fiber bundles obtained from DTI (*Diffusion Tensor Imaging*). In this work, using the *Square Root Velocity Function* (SRVF) representation proposed by Anuj Srivastava and his contributors, we aim first to build a rigorous framework for shape analysis of fibers, with the computation of summary statistics, clustering to filter out outliers and registration. Afterwards, a new perspective with *functional data analysis* is introduced, considering the functional nature of scalar measures from diffusion along fibers (i.e., *Fractional Anisotropy*). The possibility to localize abnormalities on diffusion indeces along the tract represents an innovative and challenging approach to the study of different diseases, for instance Multiple Sclerosis. We try to find a detector (significative threshold) between a patient with a known lesion in the Corticospinal tract (CST) and controls in the mean difference of the FA functions, relying on statistical inference in a functional framework. We do not claim to assess the general validity of this approach, but rather to pave a new way for future researchs.

Keywords: DTI, Shape analysis, SRVF, Fibers, FA functions, Functional Data Analysis, Multiple Sclerosis, CST.

Sommario

In un contesto di analisi quantitativa delle fibre neuronali nella materia bianca, c'è un crescente bisogno di sviluppare modelli statistici capaci di analizzare indici di diffusione lungo i principali tratti ottenuti tramite stime DTI (*Diffusion Tensor Imaging*). In questo lavoro, grazie alla rappresentazione *Square Root Velocity Function* (SRVF) proposta da Anuj Srivastava e dai suoi collaboratori, vogliamo prima modellizzare statisticamente fasci di fibre con strumenti di shape analysis, ovvero calcolo di media e varianza, clustering per eliminare outliers e registrazione. Successivamente, considerando la natura funzionale delle misure scalari di diffusione lungo le fibre (*Anisotropia Frazionaria* ad esempio), vogliamo introdurre un approccio statistico di tipo funzionale. A livello clinico, la possibilità di individuare regioni lesionate all'interno dei tratti principali della materia bianca rappresenta uno strumento efficace nell'analisi di diverse malattie, come la Sclerosi Multipla. Cerchiamo di individuare una soglia significativa nella differenza dei valori FA tra un paziente con una lesione nel tratto corticospinale e pazienti sani attraverso inferenza statistica funzionale. Le idee presentate nell'ultima parte di questa tesi, più che un approccio di validità generale, vogliono essere uno spunto verso ricerche future nella stessa direzione.

Parole Chiave: DTI, Shape analysis, SRVF, Fibre neuronali, Anisotropia frazionaria, Statistica Funzionale, Sclerosi Multipla, Tratto corticospinale.

Introduction

Background

Diffusion Magnetic Resonance Imaging (DMRI) is one of the most rapidly evolving techniques in medical imaging. This method exploits the properties of water molecules under the effect of a magnetic field. Briefly speaking, it can characterize water diffusion properties at each picture element (pixel) of an image. The key idea is to study the structure of spatial order in living organs in a non-invasively way. More precisely, in this work we wonder how the diffusion molecular displacements are constrained by the geometry of the brain.

The white matter of the **central nervous system** (CNS) constitutes the tracts in which axons extend from one CNS region to another. Within a white matter tract, the majority of axons lie parallel to one another. They may all run in one direction or distinct populations may traverse the tract in opposite directions. This oriented structure of the axonal architecture is thought to be the main reason of *anisotropic diffusion*, which means that water does not diffuse equally in all directions (for example, brain water diffuses preferentially along axonal fiber directions). Anisotropic water diffusion in neural fibers forms the basis for the utilization of DMRI techniques to track fiber pathways. The fact that water diffusion is sensitive to the underlying tissue microstructure provides a unique method of measuring the orientation and integrity of these neural fibers which may be useful in assessing a number of neurological disorders.

The most used method for identifying white matter pathways in the living human brain is MR diffusion tractography. It's quite suggestive since it's the only available tool for measuring these pathaways *non-invasively* and *in vivo*. Their non-invasive nature and ease of measurement mean that tractography studies can adress scientific and clinical questions that cannot be answered by any other means.

The main goal of this work is to perform statistical analysis on fiber bundles extracted from the tractography of a patient. Fibers can be represented such as open curves in $\mathbb{R}^3$ and an ongoing effort in medical imaging is to study the shapes of these curves to associate them with functionality of the brain. Developing statistical tools for shape analysis involves first a right mathematical representations for the shapes and second a framework for performing calculations under those representations. In particular, we will illustrate this idea using the **square-root velocity function** (SRVF), proposed by Anuj Srivastava and his contributors, which leads to an elastic shape analysis. Under this representation we aim to compute statistical summaries in different representative spaces, based on the geometry of a specific fiber bundle. Furthermore, by computing distances in this framework, we will show how one can apply these tools to cluster DTI fibers and to filter out outliers from tractography results.

After this preliminary geometric analysis, we will associate scalar measures from diffusion (such as Fractional Anisotropy) to the fiber bundle of interest. These diffusion

indeces can be viewed as functions along the fibers and they represent an efficient way to fruitfully investigate disorders in human brain. To assess these results we use tools from functional data analysis, which provide us powerful techniques for doing inference in the case of functional data.
In our case, we will try to figure out the potentiality of this approach in a Multiple Sclerosis research study, in which is known that along the course of the disease the microstructural alteration of the WM may manifest with changes in diffusion scalar measures in most of the major tracts. We now give a short view on the structure of this work, summarizing the contents of each chapter.

Organization

Chapter 1

Since Diffusion Magnetic Resonance is the basis to develop our analysis, a coincise overview on the principles of diffusion is presented in this chapter. This gives us the proper background to introduce the diffusion model used in our project, the **Diffusion Tensor Model**, with all the scalar measures derived from it and its influence in clinical application. We have choosen this model due to its flexibility and ease of application, but we underline that it has some key limitations (mainly in the detection of just one diffusion direction for each voxel). More sofisticated models can be used in order to account for more directions and thus improving the efficiency of these analysis, such as multi compartment model. Finally, tractography procedure is theoretically described with some examples.

Chapter 2

In this chapter we introduce the research study in Multiple Sclerosis and the dataset used for the analysis. Furthermore, the preprocessing of the raw data from diffusion is described in its steps.

Chapter 3

Fibers are essentially open curves with physical features such as shape, orientation, scale and position. *How can we analyze their geometry in a efficient way?* Shape analysis is the answer to this key question. In this chapter a preliminary mathematical background is given, since some concepts are fundamental to deeply understand this challenging context. Once we have introduced a comprehensive Riemaniann framework for the analysis of open curves, we face the problem of the representation that best suits our goals. The **Square Root Velocity Function**, due to its properties, is a well-reasoned choice which leads to an elastic shape analysis. Theoretical rigorous definitions of preshape and shape spaces are given in the case of shape feature, while for the joint analysis of other features only final results are listed.

Chapter 4

Here we present the application of Shape analysis in the statistical modeling of fibers. We focus on a major tract, the CST, and we model it taking into account *shape*, *orientation* and *scale*. In such a way we obtain statistical summaries and we are able to perform registration between subjects considering also a texture information along

fibers (FA in our case). We describe also a hierarchical clustering tool for filtering outliers from the bundle.

Chapter 5

In this chapter we discuss two different approaches to the problem of analyzing FA profile along the tract, the **Interval-Wise Testing** procedure and the **Functional Linear Discriminant Analysis**. The goal is to propose a method to recognize possible lesions on patients.
We consider a patient with a known lesion on the CST and first we aim to detect a significant threshold for the FA mean difference. Finally we try to validate the results.

Softwares

The Multiple Sclerosis dataset has been provided by the VISAGES team, which gave me the access to a richful online platform (SHANOIR).
About the preprocessing of the data, numerous softwares have been used: in-house MedInria and Anima for visualization of the data and registration, Freesurfer for computing the parcellation of the brain and TrackVis for the tractography.
Last, for the second part of the analysis, MATLAB and R have been largely used.

Introduzione

Background

La Risonanza Magnetica di Diffusione è una delle tecniche di maggior rilevanza nel campo dell'imaging diagnostico. Questo metodo misura la diffusione delle molecole d'acqua all'interno di tessuti biologici sotto l'effetto di un campo magnetico. L'idea chiave è quella di studiare la struttura spaziale di organi specifici in modo *non invasivo*. Nel nostro caso, ci chiediamo come gli spostamenti molecolari di diffusione sono vincolati dalla geometria del cervello umano.
La materia bianca del **sistema nervoso centrale** è costituita da fasci di fibre neuronali che connettono tra di loro diverse aree. La complicata struttura architettonica del cervello umano determina la diffusione anisotropica delle molecole d'acqua, le quali si muovono principalmente lungo le direzioni date dalle fibre. Questo fatto è alla base delle tecniche di trattografia, attraverso le quali è possibile ricostruire il percorso delle molecole per formare stime efficienti delle fibre nervose.
La trattografia è l'unica tecnica che permette di misurare questi percorsi in modo *non invasivo* e *in vivo*. La sua efficienza e la sua facilità di applicazione la rendono fondamentale a livello clinico, poichè consente di investigare la struttura celebrale e quantificare l'effetto di diverse malattie, soprattutto quelle demielinizzanti (SLA, Sclerosi Multipla).
L'obiettivo principale di questo lavoro è l'analisi statistica di tratti neurali estratti dalla trattografia del cervello di un paziente. Le fibre possono essere rappresentate come curve aperte in $\mathbb{R}^3$, caratterizzate quindi da una forma, da un orientamento e da una scala. Ci proponiamo di definire strumenti statistici adatti per lo studio di questo tipo di ogetti e lo faremo utilizzando la rappresentazione SRVF presentata da Anuj Srivastava e da suoi collaboratori. Sotto questo tipo di rappresentazione definiremo distanze in varietà non euclidee (distanze geodetiche) e calcoleremo medie e varianze, algoritmi di clustering e registrazione.
Dopo questa analisi geometrica preliminare, ci proponiamo di studiare il comportamento di indici di diffusione lungo fasci di fibre specifici. Quest'ultimi possono essere infatti visti come funzioni scalari e attraverso strumenti di analisi statistica funzionale cercheremo di fare inferenza all'interno di uno studio di ricerca in Sclerosi Multipla. Questa malattia provoca alterazioni (aree lesionate) all'interno della struttura bianca che si possono manifestare con variazioni negli indici di diffusione. Presentiamo ora la struttura di questa tesi, riassumendo i contenuti principali di ogni capitolo.

Organizzazione

Capitolo 1

Siccome la Risonanza Magnetica di Diffusione è la tecnica alla base delle nostre analisi, in questo capitolo presentiamo una sintesi dei principi fisici e teorici di questo metodo. Questo ci permette di introdurre il modello **Tensore di Diffusione** e tutte le misure scalari che ne derivano. La scelta di questo modello è dettata dalla sua flessibilità e dalle numerevoli applicazioni in campo medico. Dall'altra parte, esso è caratterizzato da alcune limitazioni chiave: è possibile ricavare solo una direzione di diffusione in ogni voxel e questa spesso è un'assunzione troppo semplicistica. I dati provenienti dall'imaging del tensore di diffusione possono essere usati per eseguire la trattografia. Presentiamo il metodo FACT e alcuni risultati a livello pratico.

Capitolo 2

In questo capitolo introduciamo brevemente le caratteristiche della Sclerosi multipla e il dataset utilizzato per le analisi. Inoltre, la parte di preprocessing dei dati di imaging è descritta nei suoi vari passaggi.

Capitolo 3

Come è stato sottolineato precedentemente, le fibre sono delle curve aperte con particolari proprietà fisiche. Come possiamo analizzare queste caratteristiche in un modo efficiente? La risposta risiede nell'approccio statistico di Shape analysis. In questo capitolo definiamo i concetti teorici fondamentali per comprendere questo contesto. La nozione di varietà riemanniana assume un ruolo primario nel modellizzare spazi curvi di dimensione arbitraria. Una volta formalizzato il giusto ambiente matematico, presentiamo la rappresentazione SRVF e gli spazi in cui svolgeremo le nostre analisi.

Capitolo 4

In questo capitolo metteremo in pratica gli strumenti descritti nel capitolo 3. Ci concentriamo su uno dei principali tratti neurali, il tratto corticospinale, considerando le caratteristiche di forma, orientamento e scala. Precisamente, andremo a calcolare una media, una struttura di varianza e cercheremo di filtrare outliers utilizzando un algorimo di clustering. Infine, includendo nell'analisi anche una misura scalare di diffusione (come l'anisotropia frazionaria), cercheremo di registrare i vari tratti ad una media comune (population mean) e di preparare il campo per l'analisi statistica funzionale.

Capitolo 5

In questo capitolo ci proponiamo di affrontare il problema della localizzazione tratto-specifica di lesioni nella sostanza bianca nel decorso della Sclerosi Multipla. Queste lesioni si possono manifestare in alterazioni di misure di diffusione lungo le fibre, come nel caso dell'anisotropia frazionaria. Quest'ultima può subire un decremento nel suo valore in aree danneggiate. Attraverso inferenza statistica funzionale cerchiamo di individuare una soglia significativa tra un paziente con lesione nel tratto corticospinale e pazienti sani.

Softwares

Il dataset è stato fornito dal team VISAGES, che mi ha concesso l'accesso ad un ricco database online (SHANOIR). Per la parte di preprocessing delle immagini, sono stati usati numerosi softwares: MedInria e Anima per la visualizzazione e registrazione, Fresurfer per il calcolo della parcellizzazione del cervello e TrackVis per la stima della trattografia. Infine, per la seconda parte delle analisi, MATLAB e R son stati largamente utilizzati.

Chapter 1

Whitte Matter Tractography

In this first chapter we present an overview of the basic physics and underlying theory of Diffusion MRI technique, followed by the introduction of the main model used for the estimation of tractography: the **Diffusion Tensor Model**.
Over the years, increasingly data acquisition scheme and techniques have been developed to better explore the complexity of diffusion in the brain [23]. The advent of Diffusion Tensor Imaging (DTI) and fiber tractography has opened new approaches in scientific understanding of many neurologic disorders. As a matter of fact, White matter tractography represents a method for identifying pathways in the living human brains *non-invasively* and *in vivo*. These pathways form the substrate for information transfer between different regions of the brain, thus its analysis is central in understanding the functionality of both normal and diseased brain. For a more detailed treatment on diffusion MR, the reader may refer to the book "Diffusion MRI: From quantitative measurement to in vivo neuroanatomy", which provides a complete review of Magnetic Resonance techniques [11].

1.1 Diffusion Magnetic Resonance

1.1.1 Principles of Diffusion

Diffusion is the movement of a fluid from an area of higher concentration to an area of lower concentration. The physical law behind this phenomenon is called Fick's first law, which relates the diffusive flux to any concentration difference through the relationship

$$\mathbf{J} = -D\nabla C$$

where **J** is the net particle flux, C is the particle concentration and D is the diffusion coefficient. As highlighted in the equation, particles move down a concentration gradient in diffusion. Another interesting feature of diffusion is that it occurs even in thermodynamic equilibrium. This is remarkable, since the Fick's law depicted above implies that when the concentration gradients or temperature vanish, there is no net flux.
The first who reports this random motion was Robert Brown (1828) in his treatment on some observations of particles trapped in cavities inside pollen grains in water. This erratic motion is best described in statistical terms by a displacement distribution.

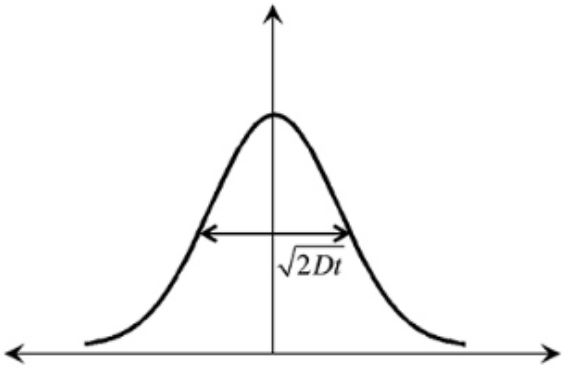


Figure 1.1: *Gaussian diffusion in one dimension (image courtesy of [11])*

This concept was formally introduced by Einstein at the beginning of the 20th century. He was able to describe the proportion of molecules that undergo displacement deriving an explicit relationship between the mean-squared displacement of the ensemble, characterizing its Brownian motion, and the classical diffusion coefficient, D, appearing in Fick's law, given by

$$\langle x^2 \rangle = 2Dt$$

where $\langle x^2 \rangle$ is the mean-squared displacement during time, Δ, and D is the same classical diffusion coefficient appearing in Fick's first law. Provided that we are in a free diffusion framework, the displacement distribution takes a Gaussian form, with the peak being at zero displacement as depicted above.
In three dimensions (if we consider for example free diffusion in a cubic volume of water) these arguments extend in a similar manner

$$\langle r^2 \rangle = 6Dt$$

Thus the isopropability surface is a sphere of radius $\sqrt{6Dt}$ and centered at the origin.
We are interested in investigating the tissue microstructure of the brain, so a natural question arises: *How can we gain information about the microstructure?*. The Einstein's equation gives us the answer, which lies in the dependence on the time of the mean squared displacement. Diffusion coefficient is then inferred from observations of the displacements over a time period and we will refer to it as the apparent diffusion coefficient, since it will be lower than the one observed in free water. Now we are ready to understand how we can actually measure displacements *non-invasively* using MRI.

1.1.2 Underpinnings of Diffusion MR

Magnetic resonance has assumed a fundamental role in clinical studies, since it provides a unique opportunity to quantify diffusional characteristics and a way to investigate the structural environment of the brain non-invasively.
Magnetic resonance images primarily reflect the signal from hydrogen nuclei, which possess a magnetic dipole, with a corresponding north and south pole. A typical MR scan starts with the excitation of the nuclei, under the main magnetic field B_0, with a 90 degree Radio Frequency pulse given by the **Larmor equation**

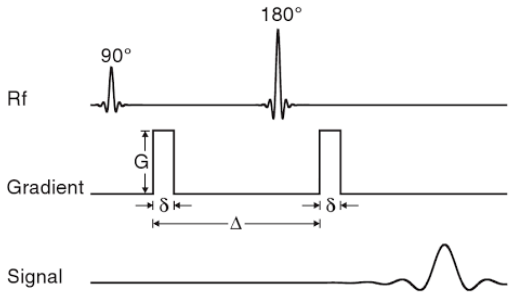


Figure 1.2: *Illustration of the* ***PGSE sequence****. δ is the duration of the short pulses separated by a diffusion time Δ and G the magnitude of the gradient (image courtesy of [11])*

$$\omega = \gamma B$$

where B is the magnetic field that the spins are exposed to and γ is the gyromagnetic ratio (for nuclei such as Hydrogeni $\gamma = 2.68 \times 10^8\, rad\, s^{-1}\, T^{-1}$). The excited spins are tilted in the plane of normal B_0, where they start precessing at the **Larmor frequency**. The transverse precessing component of the spin decays exponentially with a T_2 time constant, while the longitudinal one exponentially recovers toward its original orientation with a much slower time constant T_1.
The rotation of the precessing spins generates a magnetic field which induces a current in the receiver coils; this current is the signal used to generate MR images and is reflected in the brightness of each pixel. The measured signal will be stronger if spins are in phase, while the phase shifts will decrease it. For instance, in T_2-weighted images, contrast is produced by measuring this loss of coherence or synchrony between the spins. In a free diffusion environment, relaxation tends to take a longer time; in certain clinical situation, this can be used to detect an area of pathology.
Small inhomogeneities can occur in the magnetic field, which result in a loss of phase coherence of excited spins. A solution to this problem was proposed by Edwin Hahn (1950), who applies a second 180 degree rf, called refocusing RF pulse, at some time t. It flips the spins in the plane of normal $-B_0$, putting the faster precessing spins behind the slower ones. At time $2t$ they will be back in phase, giving form to the so-called **spin echo** principle. In this experiment, the time between the first rf pulse and the formation of the echo is called TE (*Echo Time*). The time between two successive excitation RF pulse is called TR (*Replication Time*).
To sensitize MR images to diffusion, a pulsed field gradient may be used instead of an homogeneous strong magnetic. This idea was devised by Stejskal and Tanner (1965), who introduced the basics for diffusion weighting, giving rise to the so-called **Pulse Gradient Spin Echo sequence** (see Figure 1.2).
After excitation, and before signal sampling, application of a bipolar gradient adds to each spin's precession a positive phase proportional to its average position (along the direction of the gradient) during the first gradient and a negative phase proportional to its average position during the second. The sum of these phases is related to the difference between these two positions, as it is shown in Figure 1.3
We can represent this scheme in a simple way. The net phase change induced by the first gradient pulse is given by

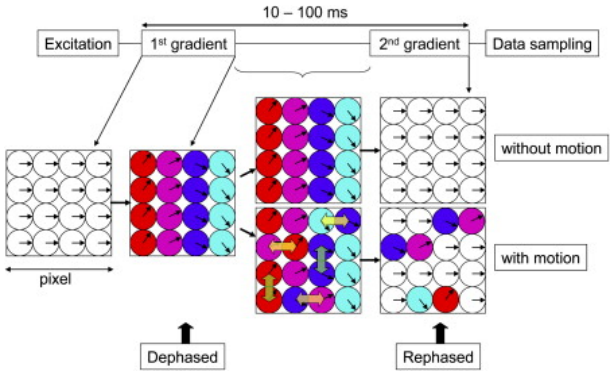


Figure 1.3: *The principle of PGSE measurements. If molecules don't undergo diffusion, there is no phase shift. (image courtesy of [35])*

$$\phi_1 = -qx_1$$

where x_1 is the position of the particle during the first pulse (note that we ignore the phase change due to B_0, since it's constant for all spins) and $q = \gamma\delta G$, where γ, δ and G are the gyromagnetic ratio, the duration and the magnitude of the gradient pulse. Similarly we have during the second gradient

$$\phi_2 = -qx_2$$

Therefore, the aggregate phase change suffered by the particle is given by

$$\phi_2 - \phi_1 = -q(x_2 - x_1)$$

if particles remained stationary, i.e. $x_1 = x_2$, the net phase shift would fade. Contrairwise, if particles diffuse randomly, the phase that they gain in the first period does not cancel the phase decrement in the second one. This incomplet cancellation causes phase dispersion and the overall signal is attenuated due to the incoherence in the orientations of individual magnetic moments. This is the fondamental relation that explains how diffusion can be measured using MR techniques.

1.1.3 Estimation of the Apparent Diffusion Coefficient

We have seen that a diffusion-weighted sequence is constructed by the addition of a pair of diffusion-sensitizing gradients, also known as motion-probing gradients. Molecular motion results in loss of signal intensity due to incomplete rephasing of water proton spins, which may change their position during the sequence application.
To generate MR contrast based on the physical properties of water molecules, proton density (PD), T_1 and T_2 relaxation times and the apparent diffusion coefficient are largely used. The proton density represents water concentration, while T_1,T_2 are related to the relaxation after excitation. This diffusion-weighted contrast can be fit to an exponential model [21, 35]:

$$S(q) = PD(1 - e^{-\frac{TR}{T_1}})e^{-\frac{TE}{T_2}}e^{-bD} = S_0 e^{-bD} \tag{1.1}$$

where TR and TE are the replication and echo time of the sequence respectively and $b = q^2 D(\Delta - \delta/3) = (\gamma\delta G)^2(\Delta - \delta/3)$ is the so-called b-value (it's used to characterize

the level of induced sensitivity on diffusion during image acquisition). Typical b values used in clinical applications range from 600 to 1500 seconds per square millimiters. Furthermore, in the equation 1.1 terms related to the costants are usually simplified as S_0.
To quantify the apparent diffusion coefficient for any directions, one needs a minimum of two signal measurements: S_0, with no diffusion-weighting gradient, and S_D, with a non-zero diffusion-weighting gradient in a desired direction. When the diffusion is isotropic, the choice of direction for the diffusion-sensitizing gradients is not relevant because the coefficient D is identical for all directions. The apparent diffusion coefficient is then estimated in the following way

$$\overline{D} = \frac{Log[S_0/S_D]}{b}$$

This formalism is extended by Stejskal's work in order to account for free anisotropic diffusion and it leads to the introduction of a general scheme, called diffusion tensor imaging (DTI).

1.2 Diffusion Tensor Model

1.2.1 Model Building

The White Matter is composed of bundles of myelinated nerve cell projections, which connect various gray matter areas. In this complicated structure we can no longer characterize the behaviour of the water molecules with a single apparent diffusion coefficient. Therefore, we must look to a more complex model for diffusion introducing a 3×3 symmetric matrix that characterizes 3D displacements and accounts for the anisotropic nature of the WM:

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}$$

For purposes of discussion let us assume that diffusion remains Gaussian but may be anisotropic. In this way we aim to model the diffusion expected along the orientation of axonal bundles. Anisotropic Gaussian distributions have six degrees of freedom instead of one. This means that, to fit our model, we must sample at least a series of six diffusion-weighted images acquired with the scheme described in the previous sections (in general with a b value of approximately $1000\, sec/mm^2$). Usual acquisition procedures account for more than six images in order to decrease the effect of the noise in the estimation. In the literature we can find different way to estimate the parameters of the tensor model [17, 12].
The diffusion tensor is usually represented by an ellipsoid, which principal axes are given by the eigenvectors and the lengths by diffusion distance in a specific time t. DT-MRI is affected also by some artifacts, such as subject motion, eddy currents, magnetic susceptibility effects and image noise [24]. Most of them are amenable to correction during postprocessing and can further be reduced with modified acquisition.

1.2.2 Scalar Measures From the Diffusion Tensor

The mathematical properties of the diffusion tensor let us to extract several useful scalar measures from diffusion tensor images. Here we present a list of the main ones,

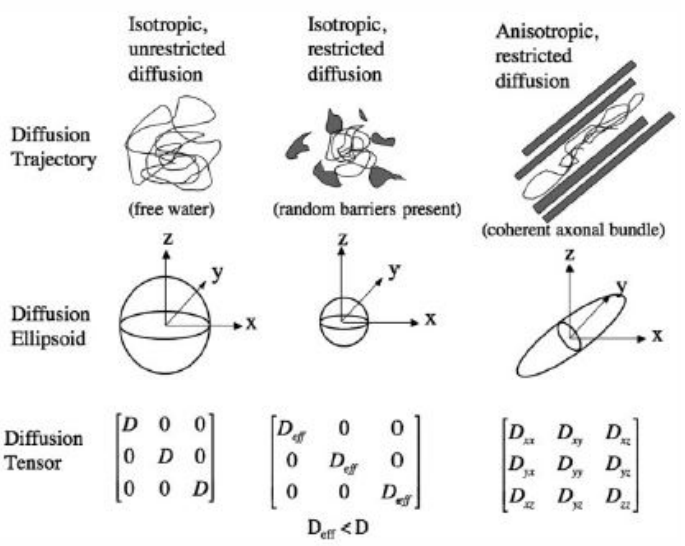


Figure 1.4: *Different possible configurations of the diffusion tensor matrix and its relative ellpsoid. (image courtesy of [21])*

which play a fundamental role in understanding in a full manner the diffusion results.

- **Trace**

  The trace is the sum of the three diagonal elements of the diffusion tensor, which is also equal to the sum of its three eigenvalues. We can think to the trace/3 as being equal to the orientationally averaged mean diffusivity. We termed it *Mean Diffusivity*

$$MD = \frac{D_{xx} + D_{yy} + D_{zz}}{3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$$

- **Fractional and Relative Anisotropy**

  Aside from describing the amount of diffusion, it is often important to quantify the relative degree of anisotropy in a voxel. A basic solution could be the ratio between the longest axis of the ellipsoid and the shortest, precisely λ_1/λ_3. However, it can be shown that this measure is very suscetible to noise. To circumvent this problem, other more complex measures are proposed, such as

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \langle\lambda\rangle)^2 + (\lambda_2 - \langle\lambda\rangle)^2 + (\lambda_3 - \langle\lambda\rangle)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}$$

  and

$$RA = \sqrt{\frac{1}{3}} \frac{\sqrt{(\lambda_1 - \langle\lambda\rangle)^2 + (\lambda_2 - \langle\lambda\rangle)^2 + (\lambda_3 - \langle\lambda\rangle)^2}}{\langle\lambda\rangle}$$

  where $\langle\lambda\rangle$ is one third of the trace of the tensor. The FA index is approximately normalized so that it is easily readable, since it takes values from zero (when

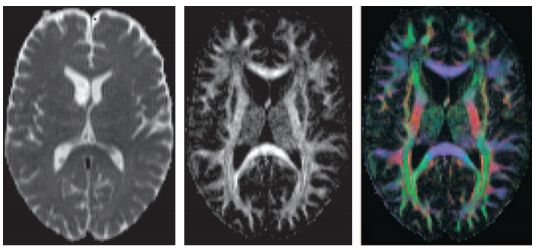


Figure 1.5: *From the left to the right: Mean diffusivity, FA index and a color-coded image showing orientations of the principal direction of diffusion. (image courtesy of [35])*

diffusion is isotropic) to one (when diffusion is constrained along one axis only). Between the two, the most used in clinical studies and in the literature is the FA index.

- **Tensor Orientation**

  By using the information contained within the diffusion tensor (mainly the eigenvector associated with the largest eigenvalue) robust fiber orientation maps could be derived. We assign color to diffusion direction, specifically red, green and blue to x, y and z components. Any other direction is assigned to a combination of these three colors, as shown in Figure 1.5.

1.2.3 Advantages and Limitations of DTI

DTI is a very popular technique in brain imaging since it provides various useful information about the tissue microstructure: it quantifies diffusion anisotropy, which is a useful measure of white matter integrity, and it provides an estimate of the principal direction of axon fiber. There are also several drawbacks. One key limitation is that it can only recover a single fiber orientation in each voxel and it fails at fiber crossing. We must also consider that although white matter may look homogenous on conventional MRI, it has a very complex structure. This makes the interpretation of the results not always straightforward [35]
We have based our analysis on this model due to its flexibility and clinical relevance in different applications, but some other more complicated models may be used [23]. The main application of DT technique in this work is the estimation of brain **tractography**, which is presented next.

1.3 Tractography

MR diffusion tractography is a method for identifying intervoxel connectivity and white matter pathways in the living human brain. DTI fiber tracking is used in clinical and scientific studies to perform localization and quantitative assessment of specifical neuronal pathways, to investigate the causes of brain diseases and to more generally analyze the functional connectivity between different areas [21, 22].
Despite his enormous potential for the study of the human brain anatomy, tractography suffers from some limitations, as it is indirect and difficult to quantify. The most crucial ones are its inability to determine the precise origin/termination of connections and to accurately track the very dense network of horizontal intracortical connections [29].

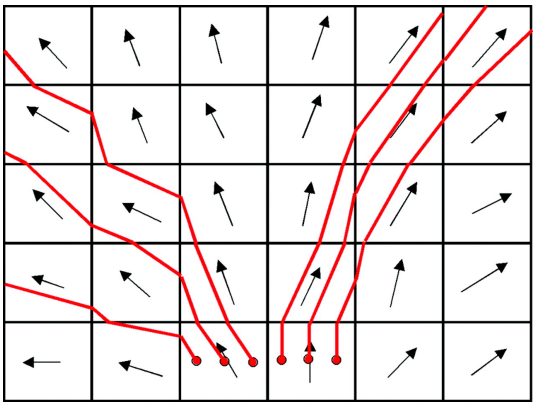


Figure 1.6: *A schematic view of the FACT algorithm. Arrow represent principal diffusion orientation and red lines the fibers tracked by the algorithm. (image courtesy of [21])*

However, advances in methods and data acquisition may help to reduce the impact of these limitations.
DTI fiber tracking algorithms can be divided into two main groups: deterministic and probabilistic methods [34, 11]. In this work we will follow the first one, using the well-known FACT algorithm.

1.3.1 Deterministic Tractography

In each brain voxel, the dominant direction of axonal tracts can be assumed to be parallel to the primary eigenvector of the diffusion tensor. How can we reconstruct fiber pathways from these principal orientations? The most intituive way to do it is considering a streamline approach. Mathematically speaking, a streamline can be viewed as a 3D space curve which location $\mathbf{r}$ is a function of the arc length s. The local principal diffusion estimation is taken to be the tangent to the streamline at the arc length s in order to estimate the fiber trajectory. In our context, we can formulate this problem as a differential equation

$$\frac{d\mathbf{r}(s)}{ds} = \mathbf{t}(s)$$

where the tangent is assumed to be the first eigenvector of the diffusion tensor, i.e. $\mathbf{t}(s) = \epsilon_1(\mathbf{r}(s))$.
Once the problem is formulated, we need a method for interpolating our discrete measurements into continous space. In the FACT algorithm, a starting seed voxel is choosen and fiber orientations are interpolated according to their nearest neighbors as depicted in Figure 1.6.
We need to be aware of the fact that any tractography process is susceptible to errors, which can have different nature:

- Imaging noise affecting the estimation of diffusion directions
- Modeling error (in our case, diffusion tensor model cannot identify regions where fiber bundles cross)
- Integration errors introduced by the algorithm

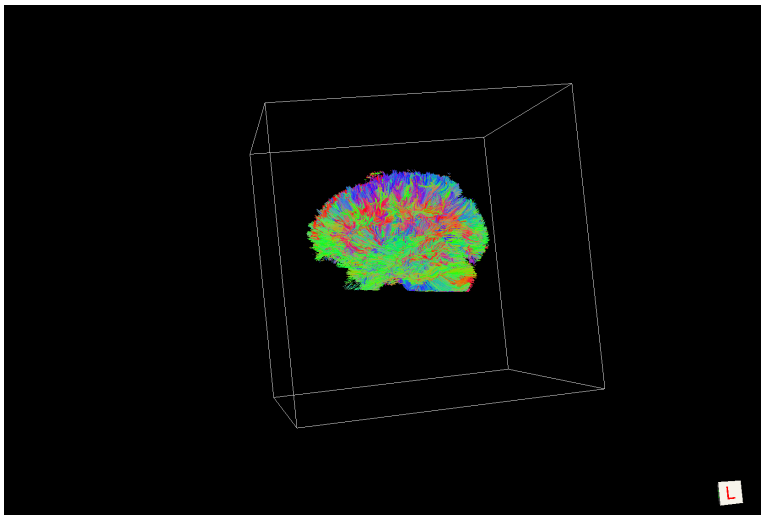


Figure 1.7: *Whole brain tractography for a single patient (scene from TrackVis)*

To overcome some of these problems, constraints on the maximum turning angle of the streamline between voxels and on the minimum FA within a voxel for propagation of the streamline can be applied. This is also a way to contain the fiber tracks to regions of the brain where the diffusion tensor model realistically represents the white matter pathways.
The choice of this basic algorithm is due to save time and effort in this part of the work, since the focus will be on the subsequent analysis of the geometry of the white matter tracts. Future directions may be on optimizing the tractography step and improving the efficiency in terms of detecting more accurate fiber tracts.

Chapter 2

Multiple Sclerosis Research Study

Multiple Sclerosis is a chronic autoimmune and infiammatory disease which affects the central nervous system. MS attacks the myelinated axons in CNS, resulting in damaged areas of varying degrees (see Figure 2.1). These damaged areas provoke difficulties to the nervous system to communicate, resulting in a range of signs and symptoms, including physical, mental, and sometimes psychiatric problems. Specific symptoms can include muscle weakness, double vision and trouble with coordination. It can take several forms and its course is higly varied and unpredictable.
While the cause is not clear, the underlying mechanism is thought to be either destruction by the immune system or failure of the myelin-producing cells and it appears to involve a combination of genetic susceptibility and a nongenetic trigger, such as a virus, metabolism, or environmental factors [18]. There is no known cure for multiple sclerosis, therefore symptomatic treatments are aimed at maintaining function and improving quality of life.
MRI plays an increasing and fundamental role in the diagnosis and management of MS [15]. It represents a reliable and accurate diagnostic technique and it is recognized as a modality able to detect and visualize lesions. These damaged areas can be characterized at MR imaging by their location, morphology and signal intensity.

2.1 Application of DTI in MS

Among all MRI techniques, Diffusion Tensor Imaging with tractography is a rich tool for providing information about the integrity of specific tract *in vivo*. By focusing on specific tract, DTI-tractography offers the possibility to investigate regional white matter damage and its impact with the disability.
In this thesis, we aim to investigate the impact of this disease in a major portion of the pyramidal sistem, the Cortico Spinal Tract (CST). It begins in the cerebral cortex, from where the neurones descend through the internal capsule, pass from the medulla and finally divides into two parts: lateral spinocortical tract and anterior spinocortical tract (see Figure 2.2). The choice of this specific tract is motivated by its linearity and robustness in the extraction procedure.
Several previous works have investigated how MS affects this tract. Associations between DTI scalar indeces and clinical measures have been investigated in [8], while in

Figure 2.1: *Lesions in red on a T_2-weigthed image of a patient affected by Multiple Sclerosis (scene from MedInria)*

[9] the attention is posed on depicting the axonal loss along the tract. From a clinical perspective, in fact, it might be expected that the effects of axonal loss would be most relevantly examined in the major tracts, such as the CST and the corpus callosum, since they play a largest role in progressive disability.
A new challenging approach is an along-tract analysis, in which the aim is to explore diffusion indeces as functions along a tract [13]. Tract profiles allow localization of focal abnormalities in subjects. There is a prominent need to consider the behaviour of these scalar measures along the entire tract, instead of considering a "tract-averaged" approach by averaging the values from the many streamline to a single estimate. We will face this topic in a new setting, using tools from functional and shape data analysis in order to compare similarities/dissimilarities between patient and control groups along scalar real-valued functions from diffusion.

2.2 Preprocessing of the data

The Multiple Sclerosis dataset is part of the USPIO-6 research study, shared in SHANOIR (Sharing NeurOImaging Resources), an open source neuroinformatics platform. The database on Shanoir is composed by $n_1 = 35$ patients and $n_2 = 20$ healthy volunteers with different ages and sex. For our analysis we have considered $n = 10$ for both the groups, since the preprocessing part takes a long time (the parcellation procedure described below may take 15-17 h per subject).
DWI data (size $128 \times 128 \times 55$, with $2mm^3$ voxels) were acquired on a $3T$ Siemens scanner. They have been denoised, but no distorsion correction has been applied.

- **Tractography Estimation**

  Once DT images are obtained, FACT tractography is computed for each patient

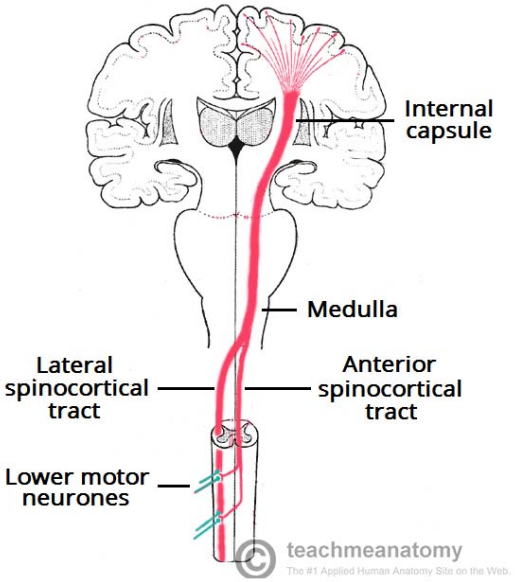


Figure 2.2: *A schematic view of the CST along the brain*

with the software **Trackvis** [28]. The stopping criteria for the tractography were $FA < 0.15$ and a threshold of 60 degrees for the curvature is used. This results in an average of tens of thousands traces per subject, as depicted in Figure 1.7. The choice of Trackvis as the software for the tractography is drived by the subsequent use of the **White Matter Query Language** (WMQL) [6], which is a novel method to extract white matter tracts from diffusion MRI volumes. It is designed in such a way to formalize tract descriptions from classic neuroanatomy textbooks and current literature on the white matter anatomy. These descriptions, formulated as a queries (with a near-to-English textual syntax), are then processed along with MRI images to extract tract of interest.

- **Parcellation**

  The WMQL requires also a parcellation of the white matter structure. This can be done using **FreeSurfer** [1] with the *recon-all* procedure, in which a full T1-image is processed in different steps. For each subject this results in the white matter parcellated as we can see in Figure 2.4. For more detailed description of this procedure, the reader may refer to the webpage `https://surfer.nmr.mgh.harvard.edu/fswiki/recon-all`.

- **Registration**

  We underline that a common registration between subjects is not necessary for our purposes. We need only to register patient-by-patient the resulted parcellation and the processed T1-image (Skull remotion and normalization) to the diffusion space in which we have computed the tractography. This can be done using

[1] Softwares FreeSurfer and FSL doesn't work on Windows Machine

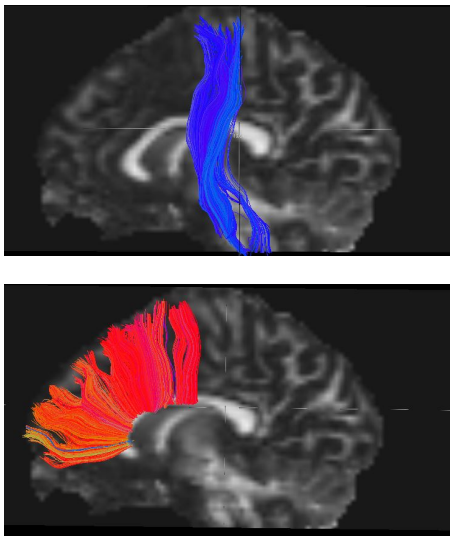


Figure 2.3: *Left Corticospinal tract and part of the Corpus Callosum extracted with WMQL plotted on a MRI volume*

different softwares, either the in-house Anima from Visages Team or FSL for linux/MacOS machine.

- **Fiber Bundle Extraction**

  Once all the above steps are done, we can run the WMQL in order to extract the fiber bundle of interest. It works from the command line and it requires in inputs the tractography file (in file format .trk or .vtk), the parcellation of the brain and the query for the specific tract. In Figure 2.3 we can see an example of a Corticospinal tract on the left hemisphere and a part of the Corpus Callosum.

Other possibilities to improve the results from the tractography may be considered, such as more complicated models for the estimation of fiber principal directions and more sofisticated fiber tracking algorithms.

2.3 Research Goals

In the current state of art, several approaches to the analysis of white matter tractography in relationship with various neurological disorders can be found. In a statistical point of view, main techniques such as **voxel-based morphometry** and **spatial statistics** have reached optimal results in the investigation of focal differences in brain anatomy.
In this work we present a different approach, trying to combine in a fruitful way geometrical notion of the fibers and scalar measures from diffusion along a bundle. There is indeed a growing need of exploring the structure of a specific white matter tract in order to quantify similarities/dissimilarities between patients and controls.
In the following chapters we will present first a preliminary geometrical analysis of the fiber bundle of interest with Shape analysis tools, motivated by the goal of building

Figure 2.4: *Parcellation of the brain obtained with FreeSurfer (scene from MedInria)*

a statistical model and performing clustering. Afterwards, a new and challenging perspective is introduced. We will try to apply methods from functional data statistic to the analysis of scalar functions (such as FA values) along fibers in order to figure out differences in damaged areas.

Chapter 3

Shape Analysis

The central theme of this thesis is the analysis of fiber tracts extracted from a brain tractography. Fibers can be represented as 3D open curves with several geometrical features, such as shape, scale, orientation and location. An ongoing effort in the medical image community is to study these fibers to investigate the functionality of the brain, since they connect one region to another and they transmit information through signals. This leads to a need of formulating a fully statistical framework in which we can do registration of points across functions and curves, quantify similarities/dissimilarities and model variability.
All these questions find an answer in a rich and powerful approach, termed **Shape Analysis**, which is becoming prominent due to the increase of datasets involving functions and curve data. There are multitudes of available techniques, which can be distinguished by the way they represent the shape itself. A very famous approach suggests to consider landmark shapes on Manifold [5, 20]. As defined in [20], a **landmark** is *"a point of correspondance on each object that matches between and within populations"*.
Salient features can then be encoded as vectors in $\mathbb{R}^{n\times d}$ where n is the number of landmarks and d the dimension of the original object. The resulting matrix of coordinates is called *configuration matrix*. Although this method had a great success, due to its innovative ideas and solutions, it does not address the issue of how to select those points to form a representative set. This kind of problem is also present in any other point-based approach.
Several recent works have introduced a flexible Riemannian framework in order to analyze shape of open curves [3, 33]. In this context curves are considered such as continous objects, despite the fact in most of the applications we have a finite number of points describing them. This approach is very richful and powerful, since in this way one does not rely anymore in the choice of representative points and registration between curves is better defined.
As a drawback, this analysis, instead of the usual one based on common vectors, brings new challenges:

- **Nonlinearity**: We will deal with nonlinear spaces, that is they are not vector spaces and one cannot use classical vector calculus to perform statistical computations.
- **Infinite Dimensionality**: Curves are formally represented by functions and spaces of functions are usually infinite dimensional.

- **Invariance**: Depending on which feature we want to analyze, we should take care of removing invariant trasformations. For example, shape is invariant to translation, rotation and rescaling.
  Another aspect of invariance is the issue of parameterization, since a curve may be parameterized in different ways without changing the shape. We will also consider this parameterization variability.

In the first sections, we will face these problems and we will briefly introduce some mathematical concepts that are relevant to understand the approach to Shape analysis. The discussion is based directly on the book [3], with some integrations on differential geometry taken from [14, 32].
Afterwards, motivated by the work of Anuj Srivastava and his contributors [3, 1, 19, 31], we will present the main representation used throughout this thesis, the **Square Root Velocity Function** (SRVF), and the different representative spaces for the analysis. In chapter 4, then, we will illustrate this framework in the context of statistical modeling of fibers and study the results with different combinations of features.

3.1 Differentiable Manifolds

Differential geometry plays an important role in shape analysis, since shape spaces are studied as differentiable manifolds and individual shapes as points on these manifolds. The commonly used tools for vector calculus (addition, multiplication etc) are not available anymore in this particular spaces.
Fibers are seen as 3D open curves, which are objects that lie in a functional space, in our case $\mathbb{L}^2([0,1],\mathbb{R}^3)$. For a ease of comprehension, we start from the idea of finite-dimensional Manifold, which is the most common situation, and then we extend it to the infinite-dimensional case.

3.1.1 Definition

Therefore, what is a Manifold? Informally speaking, a manifold is a space that locally resembles euclidean spaces, although this may not be possible globally. We now give step by step a formal definition, starting from the concept of topology.

Definition 3.1. A **topology** on a set X is a collection Ω of subsets of X, called open sets, satisfying

1. X and the empty set $\emptyset$ are in Ω.
2. The union of an arbitrary collection of open sets in Ω is in Ω.
3. The intersection of a finite collection of open sets in Ω is in Ω.

The pair (X,Ω) is called topological space. Given a topological space (X,Ω), a collection $\mathcal{P}$ of subsets of X forms a basis for the topology Ω, if

- Every set in $\mathcal{P}$ is open (i.e., $\mathcal{P} \subset \Omega$).
- For every open set $\mathcal{V} \in \Omega$ and for every $x \in \mathcal{V}$, there exists a set $\mathcal{U} \in \mathcal{P}$ such that $x \in \mathcal{U} \subset \mathcal{V}$.

One common way to obtain a topological space is to start with a metric space, which is

Definition 3.2. A **metric space** is a set X equipped with a distance function or a metric $d : X \times X \to \mathbb{R}$ with the following properties:

1. $d(x, y) \geq 0$ for all $x, y \in X$.
2. $d(x, y) = 0 \Leftrightarrow x = y$.
3. $d(x, y) = d(y, x)$ for all $x, y \in X$.
4. $d(x, y) + d(y, z) \geq d(x, z)$ for all $x, y, z \in X$ (triangle inequality).

A metric topology can be easily derived from every metric space, in which the basic open sets are open balls defined by the metric distance function:

$$B_\epsilon(x) = \{y \in X : d(x, y) < \epsilon\}$$

The sets of open $\epsilon - balls$ in X provides a basis for this topology.

Definition 3.3. A topological space is said to be **Hausdorff** if for every pair of distinct points $p,q \in X$, there exist disjoint open subsets $\mathcal{U} \subset X$ containing p and $\mathcal{V} \subset X$ containing q.

We can now give a rigorous mathematical definition of a finite-dimensional Manifold space

Definition 3.4. A topological space M is called **manifold** of dimension n if:

1. It is Hausdorff.
2. It has a countable basis.
3. For each point $p \in M$, there is a neighborhood $\mathcal{U}$ of p that is homeomorphic to an open subest of $\mathbb{R}^n$.

The *Locally Euclidean* property of M tells us that for each $p \in M$, there exists an open neighborhood $\mathcal{U}$ of p and a mapping $\phi : \mathcal{U} \to \mathbb{R}^n$ such that $\phi(\mathcal{U})$ is open in $\mathbb{R}^n$ and $\phi : \mathcal{U} \to \phi(\mathcal{U})$ is a homeomorphism (i.e., a continous bijective map [1] with continous inverse). The pair $(\mathcal{U}, \phi)$ is called a *coordinate chart* for the points that fall in $\mathcal{U}$. Note that if p denotes a point in $\mathcal{U}$, we have $(\phi_1(p), ..., \phi_n(p)) \leftrightarrow (x_1(p), ..., x_n(p))$ which are called **local coordinates** of the point p.

Definition 3.5. An *atlas* $\mathcal{A}$ for M is a collection of charts such that the images cover the whole of M:

$$M = \bigcup_{\mathcal{U} \in \mathcal{A}} \mathcal{U}$$

[1] Recall that *bijective* means both injective and surjective. The first means that $\phi(a) = \phi(b)$ implies a=b; the second means "onto", that is every element of the codomain is the image of some element of the domain.

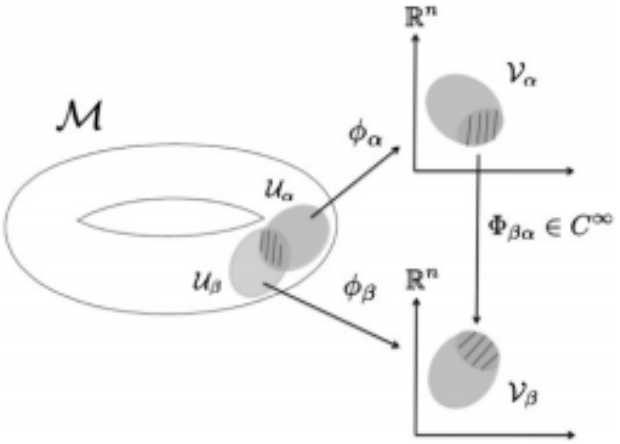


Figure 3.1: *Illustration of the application of coordinate charts* $(\mathcal{U}_\alpha, \phi_\alpha), (\mathcal{U}_\beta, \phi_\beta)$ *and its transition map* $\Phi_{\beta\alpha}$.

Two charts $(\mathcal{U}_\alpha, \phi_\alpha)$ and $(\mathcal{U}_\beta, \phi_\beta)$ are called smoothly compatible if the transition maps

$$\Phi_{\beta\alpha} = \phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \to \phi_\beta(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$$

and

$$\Phi_{\alpha\beta} = \phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \to \phi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$$

are $\mathcal{C}^\infty$, i.e. they have continous partial derivatives of all orders. (see Figure 3.1 for an illustration)

Definition 3.6. An atlas $\mathcal{A}$ is differentiable or smooth if it consists of (pairwise) compatible charts.

At this point one may observe that for a given manifold we can have many different atlases. We say that a smooth atlas $\mathcal{A}$ on M is maximal if every chart on M which is smoothly compatible with all the charts in $\mathcal{A}$ is already in $\mathcal{A}$.

Definition 3.7. M is an n-dimensional **differentiable manifold** if M is endowed with a smooth n-dimensional atlas. The number n is called the dimension of M, $n = dim(M)$.

Let's now shift the attention to the concept of differentiable maps. Consider some manifold M equipped with a smooth structure (i.e., a maximal smooth atlas $\mathcal{A}$) and a function $f : M \to \mathbb{R}$. When should it be called smooth? We define f to be smooth at $p \in M$ if $f \circ \phi^{-1}$ is smooth for all charts in the atlas. The set of all such smooth real-valued functions at p is denoted by $C^\infty(p)$.
In a similar manner we can define smoth maps between two manifolds $F : M \to N$. For any point $p \in M$, consider local charts $(\mathcal{U}, \phi)$ with $p \in \mathcal{U}$ and $(\mathcal{V}, \psi)$ with $F(p) \in \mathcal{V}$. If the mapping $\psi \circ F \circ \phi^{-1} : \phi(\mathcal{U} \cap F^{-1}(\mathcal{V})) \to \psi(\mathcal{V})$ is smooth at the point $\phi(p)$ (in the sense that all its partial derivatives exist and are continous) then F is called **smooth** at p. A pictorial illustration is shown in Figure 3.2. Again, beacuse of the smooth compatibility of charts, this determination will not depend on the choice of them.

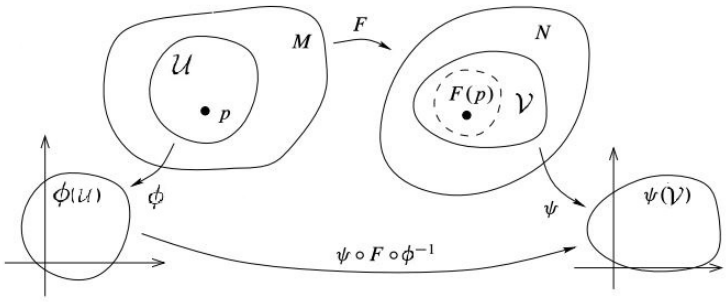


Figure 3.2: *Illustration of the smooth map $F : M \to N$ (image courtesy of [14])*

3.1.2 Tangent space

The advantage about a differentiable manifold M is that we can evaluate derivatives of functions on M in terms of any coordinate chart we wish, and if we change them, these derivatives will be related by the chain rule. Once we have the definition of a differentiable manifold, we want to perform differential calculus of functions on these spaces and to do this one needs to introduce the concept of the tangent structure.
There are various equivalent ways of defining the tangent spaces of a manifold. We first face this argument considering differentiable (C^1) curves on the manifold passing through the point of interest and studying the velocity vectors of these curves at that point.
A tangent vector at a point $p \in M$ is thought as the "velocity" of a curve passing through the point p. We can therefore consider an equivalence class of curves passing through p, which elements are the tangent vectors at p.
More formally, let M be a n-manifold and for a point $p \in M$ we consider a curve $\gamma : (-\epsilon, \epsilon) \to M$ such that $\gamma(0) = p$. We then pick a chart $(\mathcal{U}, \phi)$, where $\mathcal{U}$ is an open subset containing p. Since γ is differentiable and ϕ smooth, the composition $\mu = (\phi \circ \gamma) : (-\epsilon, \epsilon) \to \mathbb{R}^n$ is differentiable as well. Its derivative denotes the velocity of γ at p in local coordinates.
Any two differentiable curves γ_1, γ_2 passing through p at $t = 0$ are called equivalent if the ordinary derivatives of μ_1, μ_2 coincide at 0, i.e.

$$\frac{d}{dt}\mu_1(0) = \frac{d}{dt}\mu_2(0)$$

This defines an equivalence relation and the equivalence classes are known as the tangent vectors of M at p. They take the form

$$[\gamma] = \left\{ \beta : (-\epsilon, \epsilon) \to M | \beta(0) = p \ \ and \ \ \frac{d}{dt}(\phi \circ \beta)(t)_{|t=0} = \frac{d}{dt}(\phi \circ \gamma)(t)_{|t=0} \right\}$$

A tangent vector to M at p is defined to be one of these equivalence classes and the set of all such tangent vectors is called the tangent space to M at p, or $\mathcal{T}_p(M)$.
To define the operations on $\mathcal{T}_p(M)$ we use a bijective map $\mathcal{T}_p(M) \to \mathbb{R}^n$, i.e $[\gamma] \to \frac{d}{dt}(\phi \circ \gamma)(t)_{|t=0}$. It turns out that this map can be used to transfer the vector space operations from $\mathbb{R}^n$ to $\mathcal{T}_p(M)$, turning the latter into an n-dimensional real vector space. We underline also the fact that the choice of the chart doesn't affect the above costruction.

The bijective map above is related with the concept of **derivations**. Let $C^\infty(p)$ be the set of all smooth real-valued functions as defined earlier. For a point $p \in M$, a derivation is a linear map $X : C^\infty(p) \to \mathbb{R}$ with a property modeled on the product rule of calculus

$$X(fg) = X(f)g(p) + X(g)f(p), \qquad f, g \in C^\infty(p)$$

If we define addition and scalar multiplication for such derivations we get a real vector space which is defined to be the tangent space and there is a one-to-one correspondence between tangent vectors and derivations. In the case of $M = \mathbb{R}^n$, the tangent space is $T_p(\mathbb{R}^n) = \mathbb{R}^n$ and we can define the mapping $v \to X_v \equiv \sum_{i=1}^n v_i \frac{\partial}{\partial x_i}$, where $\frac{\partial}{\partial x_i}$ are example of derivations. This association underlines also the usage of tangent vectors for computing **directional derivatives** [2]. For instance, for a function $f \in C^\infty(p)$, the notation $vf(p)$ stands for the directional derivative of f in the v direction and is defined to be $vf(p) = \frac{d}{dt}(f \circ \gamma(t))|_{t=0} \equiv X_v(p)$. Thus, one can interchangeably consider a tangent vector as a derivative operator for functions at p or velocity vector of a curve passing through p.
These notions lets us to define a differential of a mapping between two manifolds

Definition 3.8. The differential of a smooth mapping $f : M \to N$ at $p \in M$ is a linear map $df_p : T_p(M) \to T_{f(p)}(N)$ specified as follows. Let $g : N \to \mathbb{R}$ be a smooth function. Then, for any $v \in T_p(M)$, define

$$(df_p(v))(g) = v(f \circ g)(p) \equiv X_v(f \circ g)$$

Looking at the right side, the derivative is given by forming a new function $f \circ g : M \to \mathbb{R}$ and computing its directional derivative at p using v. In other words, the directional derivative of g on N is defined as the directional derivative of the composition with f on M.
Along the analysis, most of the sets we will encounter will be treated as susbsets of larger manifolds. There is indeed a need of defining the concept of **submanifold**. A useful way to obtain them is to exploit the definition above. A point $p \in M$ is said to be a **critical point** of f if the differential $df(p)$ is not onto, while a **regular point** if it is. The image of a critical point, $f(p)$, is called a **critical value** and any point $q \in N$ which is not critical is a **regular value**. The following holds:

Theorem 3.1. *Suppose M and N to be manifolds of dimension m and n respectively and let $f : M \to N$ be a smooth map with a regular value $y \in N$. Then $f^{-1}(y)$ is a submanifold of M with dimension m-n.*

Example 3.1. An important example of differentiable manifold for our case is the **Unit Sphere** in $\mathbb{R}^{n+1}$, defined as follows:

$$\mathbb{S}^n = \left\{ p \in \mathbb{R}^{n+1} | \sum_{i=1}^{n+1} p_i^2 = 1 \right\}$$

Let's now check indeed if it's a submanifold of $\mathbb{R}^{n+1}$. We define $f : \mathbb{R}^{n+1} \to \mathbb{R}$ to be a map given by $f(p) = \sum_{i=1}^{n+1} p_i^2$. Its differentialf $df_p(u) = 2\langle p, u \rangle$ is clearly onto for all $p \in f_{-1}(1)$. Thus, 1 is a regular value of f and the set $f_{-1}(1)$ given by $\mathbb{S}^n$ is a n-dimensional submanifold of $\mathbb{R}^{n+1}$.

[2] Generally, the *directional derivative* of a differentiable function along a given vector v at a given point p intuitively represents the instantaneous rate of change of the function.

3.1.3 Extension to the Infinite-dimensional Case

This theory can be extended to an infinite-dimensional setting, in which a manifold is modeled on Hilbert spaces. This leads to the concept of Hilbert manifolds. In the context of shape analysis we are interested in this spaces, since one aims to analyze curves which are elements of $\mathbb{L}^2([0,1],\mathbb{R}^n)$. The latter is in fact a Hilbert space, i.e. a Banach space in which the norm is defined in terms of an inner product: $||\cdot|| = \sqrt{\langle\cdot,\cdot\rangle}$, where

$$\langle f_1, f_2\rangle = \int_0^1 f_1(x)f_2(x)dx, \qquad f_1, f_2 \in \mathbb{L}^2([0,1],\mathbb{R}^n)$$

We repeat here the formal definition of smooth manifold, which takes now a more general form.
First we define the notion of a smooth atlas:

Definition 3.9. Let X be a topological space. A **smooth atlas** on X is a collection of pairs (U_i, ϕ_i) satisfying the following conditions:

1. Each U_i is an open subset of X and the $X = \bigcup U_i$.
2. Each ϕ_i is a homeomorphism of U_i onto the open set $\phi_i(U_i)$ of some Banach space E_i and for any i,j the set $\phi_i(U_i \cap U_j)$ is open in E_i.
3. The transition map:

$$\phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j)$$

is a smooth isomorphism [3] for each pair i, j.

If we are given a new chart (U, ϕ), we shall say that it is **compatible** with the atlas (U_i, ϕ_i) if each transition map $\phi_i \circ \phi^{-1}$ is a smooth isomorphism. Two atlases are said to be **compatible** if each chart of one is compatible with the other atlas. We can now give the following

Definition 3.10. (**Smooth Manifold**) A topological space X with a choice of an equivalence class of smooth atlases is called a **smooth manifold**.

Note that if $E = \mathbb{R}^n$ for some fixed n, then we have the previous case of finite-dimensional manifold.
Finally, for infinite-dimensional submanifolds a similar result to the finite-dimensional case presented earlier holds. Let X and Y be two smooth Banach manifolds and let $f : X \to Y$ be a smooth mapping. The map f is called transversal over y if the differential of f at every point of $f^{-1}(y)$ is onto.

Theorem 3.2. *If f is transversal over y, then $f^{-1}(y)$ is a **submanifold** of X.*

As we will see in the following sections, a well known manifold used througout our analysis is the unit radius hypershpere. Its relative simple geometry will let us to derive tools for performing shape analysis.

[3] An isomorphism of smooth manifolds is commonly said *diffeomorphism*. It is an invertible function that maps one differentiable manifold to another such that both the function and its inverse are smooth.

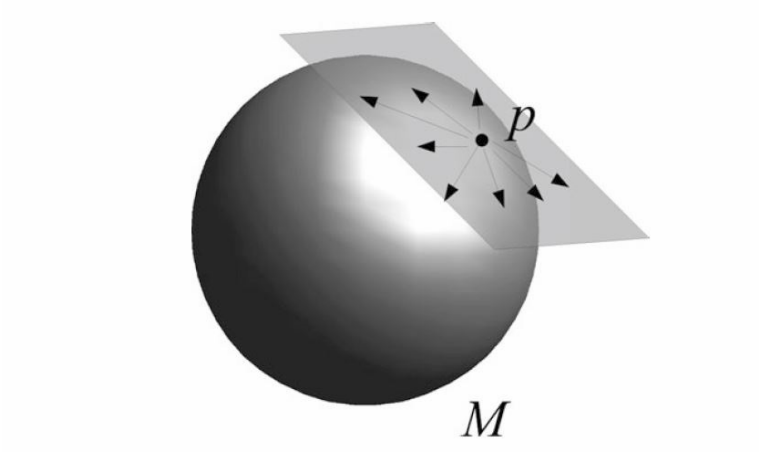


Figure 3.3: *The tangent space at a point p on a hypersphere, which is a submanifold of* $\mathbb{L}^2([0,1])$.

Example 3.2. Let $\mathbb{L}^2([0,1])$ be the infinite-dimensional space and $f : \mathbb{L}^2([0,1]) \to \mathbb{R}$ defined by $f(g) = \langle g, g \rangle = ||g||$ the transversal map. Then:

$$\mathbb{S}_\infty = \left\{ g \in \mathbb{L}^2([0,1]) | ||g|| = 1 \right\}$$

is a submanifold of $\mathbb{L}^2([0,1])$. We call it **hypersphere**. For any $g \in \mathbb{S}_\infty$, the tangent space $T_g(\mathbb{S}_\infty)$ is given by:

$$T_g(\mathbb{S}_\infty) = \left\{ h \in \mathbb{L}^2([0,1]) | \langle h, g \rangle = 0 \right\}$$

In Figure 3.3 is shown an hypersphere with an example of tangent space.

Now we have sufficient mathematical knowledge to introduce a particular kind of manifold space, the **Riemannian manifold**. This will be the right space to build on all the statistical tools for calculating distances, summary statistics and probabibility models.

3.2 Riemannian Framework

3.2.1 Riemannian Metric and Geodesics

Based again on [3], we are going to introduce a flexible Riemannian framework for our analysis. How can we compute distances between curves on a manifold? This is a key question, since distances on shape manifolds can be used to quantify dissimilarities/similarities between shapes. In this way we may characterize fibers in different bundles and in different patients.
These distances will be calculated by constructing *shortest paths* between shapes and by measuring the lengths of these paths. First we need to define a particular kind of metric, the so-called **Riemannian metric**.

Definition 3.11. A **Riemannian metric** on a differentiable manifold M is a map ϕ that smoothly associates to each point $p \in M$ a symmetric, bilinear, positive definite form on the tangent space $\mathcal{T}_p(M)$, i.e.

$$\phi : \mathcal{T}_p(M) \times \mathcal{T}_p(M) \to M, \qquad p \in M$$

A differentiable manifold endowed of this Riemannian metric is called **Riemannian manifold**.

Example 3.3. For the unit sphere $\mathbb{S}^n$ and a point $p \in \mathbb{S}^n$, the Euclidean inner product on the tangent space make $\mathbb{S}^n$ a Riemannian manifold, that is for $v_1, v_2 \in \mathcal{T}_p(\mathbb{S}^n)$ we use the Riemannian metric $\phi(v_1, v_2) = {v_1}^T v_2$.

This Riemannian structure let us to calculate lengths of paths on a manifold. The length is simply given by the following integral

$$\mathcal{L}[\alpha] = \int_0^1 \sqrt{(\phi(\frac{d\alpha(t)}{dt}, \frac{d\alpha(t)}{dt}))}$$

where

- $\alpha : [0,1] \to M$ is a parameterized path on M,
- $\frac{d\alpha(t)}{dt}$ is a velocity vector on the tangent space $\mathcal{T}_p(M)$,
- ϕ is the metric defined previously.

For any two points $p,q \in M$ we can define the distance between them as the infimum of the lengths of all smooth paths on M that start at p and end at q:

$$d(p,q) = \inf_{\alpha:[0,1]\to M | \alpha(0)=p, \alpha(1)=q} \mathcal{L}[\alpha] \tag{3.1}$$

With this definition of distance, M becomes a metric space and the metric topology agrees with the manifold one. Hence, geodesics in a Riemannian manifold are simply the locally distance-minimizing paths.

Definition 3.12. If there exists a path $\hat{\alpha}$ that achieves the minimum in equation 3.1, then it is called a geodesic between p and q on M.

Example 3.4.

1. Geodesics on a unit hypersphere $\mathbb{S}^n$ are great circles. The minimizing one is the shorter of the two arcs joining any two points. Precisely, for any p,q we have:

$$\alpha(\tau) = \frac{1}{\sin(\tau)}(\sin(\vartheta(1-t)p + \sin(\vartheta t)q)$$

where ϑ is determined by $\cos(\vartheta) = \langle p, q \rangle$ and $0 < \vartheta < \pi$

2. Geodesics on $\mathbb{R}^n$ with the Euclidean metric are straight lines, i.e. for $p, q, \in \mathbb{R}^n$

$$\alpha(\tau) = \tau q + (1-\tau)p$$

Another important concept is the exponential map, which lets us to transfer back and forth between M and $\mathcal{T}_p(M)$. It is of great interest, since the vector space $\mathcal{T}_p(M)$ can be viewed as a locally flat approximation of M and one can do multivariate analysis on it. Before giving the definition, we need to enunciate the following

Theorem 3.3. *Let M be a Riemannian manifold. Given a point $p \in M$ and a tangent vector $v \in \mathcal{T}_p(M)$, there exixsts a unique geodesic $\alpha_v : (-\epsilon, \epsilon) \to M$, for some $\epsilon > 0$, such that $\alpha_v(0) = p$ and $\dot{\alpha_v}(0) = v$.*

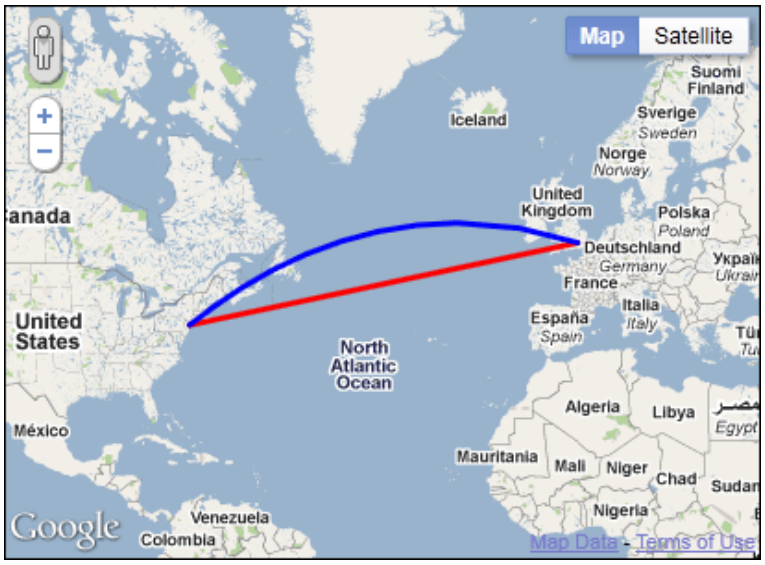


Figure 3.4: *The Earth could be viewed as a hypersphere. The shortest path between two points is then the great circle in blue (image courtesy of Gabriel Svennerberg, 2011).*

Based on the theorem above, the definition of the exponential map follows straightforwardly

Definition 3.13. Let p be a point of a Riemannian manifold M. The exponential map $exp_p : U \subset \mathcal{T}_p(M) \to M$ is defined in the following way:

$$exp_p(v) = \alpha_v(1)$$

where α_v is as defined in the theorem above and U contains an open neighborhood of the origin in $\mathcal{T}_p(M)$.

There exists also the inverse of an exponential map, which takes a point on the manifold M and maps it to an element (or multiple elements) of the tangent space $\mathcal{T}_p(M)$.

Example 3.5. The geodesics on a hypersphere $\mathbb{S}^n$ under the Euclidean metric can also be parameterized in terms of a direction v in $\mathcal{T}_p(M)$:

$$\alpha_t(v) = \cos(t|v|)p + \sin(t|v|)\frac{v}{|v|}$$

Thus, the exponential map can be defined as:

$$exp_p(v) = \cos(|v|)p + \sin(|v|)\frac{v}{|v|}$$

For a point $q \in \mathbb{S}^n$ $(q \neq p)$, the inverse exponential map $exp_p^{-1}(q)$ is given by u, where:

$$u = \frac{\theta}{\sin(\theta)}(q - \cos(\theta)p), \qquad \theta = \cos^{-1}(\langle p, q \rangle)$$

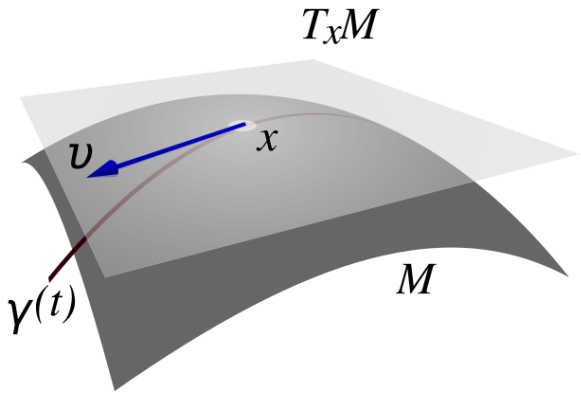


Figure 3.5: *Exponential map associates to the vector v the unique geodesic passing through it (image courtesy of McSush, wikipedia).*

3.2.2 Group Actions and Quotient Space

A central point of shape analysis is the use of transformation to model variations between objects, especially within the same shape classes. These transformations are formulated as actions of particular groups, as follows

Definition 3.14. (**Lie Groups**) A group G is a **Lie group** if

- it is a smooth manifold
- the group operations $G \times G \to G$ defined by $(g, h) \to gh$ and $G \to G$ defined by $g \to g^{-1}$ are both smooth mappings.

Our goal is to study how the application of a group changes points on a manifold M, since in our shape analysis we are interested to remove certain kinds of transformations from the representation.

Definition 3.15. Given a manifold M and a Lie group G, a left group action of G on M is a map $G \times M \to M$, given by $(g, p) \to g * p$ such that:

1. $g_1 * (g_2 * p) = (g_1 \cdot g_2)p$, $\forall g_1, g_2 \in G$ and $p \in M$.
2. $e \cdot p = p$, $\forall p \in M$.

We can also formulate the *right* version in a similar way. In case M has a Riemannian structure with a proper distance on it, we call the group action **isometric** if it preserves the Riemannian metric on M.

Example 3.6. The groups related to shape-preserving transformations are Lie Groups:

1. The translation group $\mathbb{R}^n$ is both a manifold and a group. Therefore is a Lie Group. Its action is given by $x \cdot y = x + y, \ \ \forall x, y \in \mathbb{R}^n$.
2. The scaling group $\mathbb{R}^\times$ is a Lie Group. It acts on $\mathbb{R}^n$ by $a \cdot x = ax \ \ \forall a \in \mathbb{R}^\times$ and $x \in \mathbb{R}^n$.

In case there are multiple groups acting on a manifold, we call it a direct product action. Let G and H be two groups, we give to the product $G \times H$ the structure $(g_1, h_1) \cdot (g_2, h_2) = (g_1 \cdot g_2, h_1 \cdot h_2)$. We say these actions commute if $g*(h*p) = h*(g*p)$ for all $g \in G$, $h \in H$ and $p \in M$. In the case that it is verified, they combine to give us an action of the product group $G \times H$ on m defined by $(g, h) * p = g * (h * p)$.

Example 3.7. *Scale and Rotate*: Consider the rotation group $SO(n)$ and the scaling group $\mathbb{R}^\times$. We can form the product group $SO(n) \times \mathbb{R}^\times$ and for $x \in \mathbb{R}^n, O \in SO(n)$ and $a \in \mathbb{R}^\times$ the action do commute, i.e $a(Ox) = O(ax)$.

We will use group actions to incorporate the role of shape-preserving transformations in shape analysis. There is in fact an additional notation associated with the action of a group

Definition 3.16. Consider a group G acting on a manifold M. The orbit of an element $p \in M$ is the set of elements in M to which p can be moved by the elements of G It is denoted by $G \cdot p = \{g \cdot p | g \in G\} \equiv [p]$

Finally we introduce the notion of quotient spaces of Riemannian manifolds

Definition 3.17. Let M be a manifold and G a Lie group that acts on M. Then we can define $M/G = \{[p] | p \in M\}$ to be the quotient space.

We now detail the application of these notions and techniques in shape analysis, where we aim to impose a certain Riemannian structure on different representative spaces and to describe the computation of geodesic paths and distances on them.

3.3 Representation of Curves

So far we have discussed about the main challenges that one have to face in a shape analysis framework. In this section we will develop solutions to the problems of shape-invariant trasformations, non linearity and infinite dimension considering the *square root velocity function* (SRVF).
We will carry out the general procedure in the case of shape space under the SRVF framework, which enables us to obtain tools for different tasks:

- computation of geodesic distances, geodesic paths and optimal registrations between curves.
- summary statistics such as mean, covariance and PCA.
- clustering among curves

All these topics can be straightforwardly altered to include different features in the analysis. Each representation space we will define uses a unique combination of features and, by extension, invariances but they share common procedural steps.

3.3.1 Definition and Motivation

Let $\beta : [0, 1] \to \mathbb{R}^3$ be an absolutely continuous curve arbitrarily parameterized. We introduce the following:

Definition 3.18. Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ to be a map given by:

$$F(v) = \begin{cases} \frac{v}{\sqrt{(|v|)}}, & \text{if } |v| \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

Then, we define SRVF to be a function $q : [0,1] \to \mathbb{R}^3$, where $q(t) = F(\dot{\beta}(t)) = \frac{\dot{\beta}(t)}{\sqrt{|\dot{\beta}(t)|}}$ and $|\cdot|$ is the 2-norm in $\mathbb{R}^3$.

One may immediately note that the SRVF is well defined, since it's a continous map and it exists also for curves with singular representations. In addition, if a curve β in a range $[a,b]$ is of length l, then we have $\int_a^b |q(t)|^2 dt = \int_a^b |\dot{\beta}(t)|^2 = l$. One can also recover the curve from q, up to a translation, using $\int_a^b |q(\tau)|q(\tau)dt$.
A question arises: *What makes this representation so special?* There are several reasons for selecting it [30, 31]. First, the use of the SRVF leads to a complete and powerful elastic analysis of curves, where we consider trasformations of curves that result not just from bending but from stretching as well. The **elastic metric** quantify this amount of deformation. In [3], Anuj Srivastava and his contributors highlight the relationship between the SRVF representation and the elastic metric.
Let $\beta : [0,1] \to \mathbb{R}^3$ be a smooth curve and $\dot{\beta}(t) \neq 0$ for all t. We introduce:

$$\phi(t) = \ln(\dot{\beta}(t)), \qquad \Phi = \{\phi : [0,1] \to \mathbb{R}\}$$
$$\theta(t) = \frac{\dot{\beta}(t)}{|\dot{\beta}(t)|}, \qquad \Theta = \left\{\theta : [0,1] \to \mathbb{S}^{n-1}\right\}$$

where $\phi(t)$ is the instantaneous log of the speed and $\theta(t)$ is the instantaneous direction of the curve. These two functions determine unequivocally the curve, since $\dot{\beta}(t) = e^{\phi(t)}\theta(t)$. In this way we have built a new map from the space of parameterized curves in R^3 to $\Phi \times \Theta$.
The tangent space of $\Phi \times \Theta$ at any point (ϕ, θ) is defined as $\mathcal{T}_{(\phi,\theta)}(\Phi \times \Theta) = \{(u,v) : u \in \Phi \ \textit{and} \ v : [0,1] \to \mathbb{R}^n \ \textit{smooth and} \ v(t) \perp \theta(t), \forall t \in [0,1]\}$

Definition 3.19. For any point $(\phi, \theta) \in (\Phi, \Theta)$, a pair (u_1, v_1) and (u_2, v_2) in $\mathcal{T}_{(\phi,\theta)}(\Phi \times \Theta)$, the **elastic metric** is defined as the inner product by

$$\langle (u_1, v_1), (u_2, v_2) \rangle_{(\phi,\theta)} = a^2 \int_0^1 u_1(t)u_2(t)e^{\phi(t)}dt + b^2 \int_0^1 \langle v_(t), v_2(t) \rangle e^{\phi(t)}dt \quad (3.2)$$

The two integrals in (3.2) account for the amount of stretching and bending, respectively, weighted by the constants *a,b*. We can relate the SRVF with (ϕ, θ) representation, i.e. $q(t) = e^{\frac{1}{2}\phi(t)}\theta(t)$. The following theorem proves the importance of the q-representation of curves

Theorem 3.4. *The $\mathbb{L}^2$ metric on the space of square root velocity functions for curves in R^3 correspond to the elastic metric on $\Phi \times \Theta$ with $a = \frac{1}{2}$ and $b = 1$.*

This is a remarkable result: the $\mathbb{L}^2$ metric on the SRVF space gives us an elastic framework and, as we will see in the next section, the space of fixed-length curves is a unit hypersphere under it, with a well-known geometry.

Last but not least, the reparameterization group, under the metric defines above, acts by isometries on the representation space of SRVFs, that is for q_1, q_2 and $\gamma : [0,1] \to [0,1]$ we have $||q_1 - q_2|| = ||(q_1, \gamma) - (q_2, \gamma)||$. This property will be useful when we'll treat the reparameterization variability on the shape space.

3.3.2 Preshape and Shape spaces under SRVF

We now have selected the representation for our parameterized curves in $\mathbb{R}^3$: the elastic curve representation (i.e., the SRVF) under the $\mathbb{L}^2$ metric. In the context of shape analysis, the next step is to make it invariant to all the shape-preserving trasformations (translation, rotation, scaling) and also to the reparameterization variability.

1. **Preshape Space**

   The translation variability is already removed since we consider the velocity function, while the scaling variability can be easily removed defining the space of fixed-length curves in the following way:

   $$\mathcal{C}_1 = \left\{ q : [0,1] \to \mathbb{R}^3 | \int_0^1 |q(t)|^2 dt = 1 \right\}$$

   By definition, this is a hypersphere of unit radius in $\mathbb{L}^2([0,1], \mathbb{R}^n)$. We can define a Riemannian metric:

   $$\langle w_1, w_2 \rangle = \int_0^1 \langle w_1(t), w_2(t) \rangle dt$$

   where the inner product in the integrand is the standard Euclidean product between vectors. For any $q \in \mathcal{C}_1$, the tangent space is simply:

   $$\mathcal{T}_q(\mathcal{C}_1) = \left\{ w : [0,1] \to \mathbb{R}^3 | \langle w, q \rangle = 0 \right\}$$

   The differentiable manifold $\mathcal{C}_1$ with this structure becomes a *Riemaniian Manifold*. We underline that the elements of $\mathcal{C}_1$ don't represent the shape of a curve uniquely. We have not accounted yet for all the variability generated by shape-preserving transformations.

2. **Shape space**

   A reparametrization of the curve, using a diffeomorphism $\gamma \in \Gamma_1 : [0,1] \to [0,1]$, results in a different SRVF while preserving the shape. Suppose we have β_1, β_2 such that $\beta_2 = \beta_1 \circ \gamma$, what is the relationship between q_1 and q_2? Using the chain rule:

   $$q_2(t) = \frac{\beta_1(\dot{\gamma}(t))\dot{\gamma(t)}}{\sqrt{|\beta_1(\dot{\gamma}(t))\dot{\gamma(t)}|}} = \frac{\beta_1(\dot{\gamma}(t))}{\sqrt{\beta_1(\dot{\gamma}(t))}} \sqrt{\dot{\gamma(t)}} = q_1(\gamma(t)) \sqrt{\dot{\gamma(t)}}$$

   Additionally, any rigid rotation of β changes q but non its shape. The action of the rotation group $SO(3)$ is given by $(O, q) = Oq$. Two fundamental results hold (proved in [1]):

Lemma 3.1. *The actions of* Γ_1 *and* $SO(3)$ *on* $\mathbb{L}^2$ *commute*

Lemma 3.2. *The actions of* Γ_1 *and* $SO(3)$ *on* $\mathbb{L}^2$ *are by isometries, i.e.* $||O(q_1,\gamma)-O(q_2,\gamma)|| = ||q_1 - q_2||$ *for all* $q_1, q_2 \in S_1$, $\gamma \in \Gamma_1$ *and* $O \in SO(3)$.

To unify all the different representations of the same curve, we can now define an *equivalence class* or *orbit* [4] of functions:

$$[q] = closure\left\{O\sqrt{\dot{\gamma}}(q\circ\gamma)|\gamma\in\Gamma_1, O\in SO(3)\right\}$$

The set of all these orbits is defined as the shape space, which is a quotient space: $\mathcal{S}_1 = \mathcal{C}_1/(\Gamma_1 \times SO(3))$. Since the shape space is a quotient of the preshape, it inherits the same Riemannian metric. With this structure one can compute geodesic distances.

3.3.3 Geodesics in the representative space

With the Riemannian structure and the spherical geometry of the preshape space, we can derive an analytical form for geodesics between curves: they are given by the shorter arcs on great circles

$$\alpha(\tau) = \frac{1}{\sin(\theta)}(\sin(\theta(1-\tau))q_1 + \sin(\tau\theta)q_2)$$

where $\theta = d_{\mathcal{C}_1}(q_1,q_2) = \cos^{-1}(\langle q_1, q_2\rangle)$. This geodesic starts at q_1 at $\tau = 0$ and reaches q_2 at $\tau = 1$, while traveling at a constant speed. According to example 3.5, we can define the exponential map $exp : T_q(\mathcal{C}_1) \to \mathcal{C}_1$ as:

$$exp_q(w) = \cos(||w||)q + \sin(||w||)\frac{w}{||w||}$$

Furthermore, for any $q_2 \in \mathcal{C}_1$, the inverse of the exponential map at $q_1 \in \mathcal{C}_1$, which is denoted by $exp_{q_1}^{-1} : \mathcal{C}_1 \to T_{q_1}(\mathcal{C}_1)$, is computed as follows:

$$exp_{q_1}^{-1}(q_2) = \frac{\theta}{\sin(\theta)}(q_2 - \cos(\theta)q_1)$$

These are all the basic ingredients for characterizing curves in the preshape space and they can be extended in a similar manner to the shape space. The distance is inherited and is formulated in the following way:

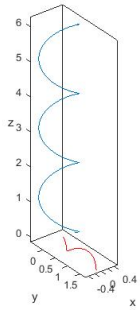
$$d_{\mathcal{S}_1}([q_1],[q_2]) = \inf_{\gamma\in\Gamma_1, O\in SO(3)} d_{\mathcal{C}_1}(q_1, O\sqrt{\dot{\gamma}}(q_2\circ\gamma))$$

This distance can be written in a more fascinated and suitable form considering the $\mathbb{L}^2$ norm, which opens up the possibility of computationally efficient solutions:

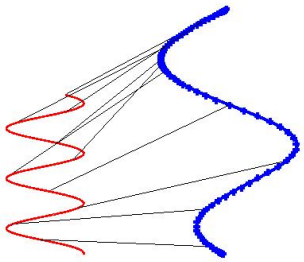
$$\inf_{\gamma\in\Gamma_1, O\in SO(3)} \cos^{-1}(\langle q_1, O\sqrt{\dot{\gamma}}(q_2\circ\gamma)) = \operatorname*{arginf}_{\gamma\in\Gamma_1, O\in SO(3)} \left\| q_1 - \sqrt{\dot{\gamma}}O(q_2\circ\gamma)\right\|^2 \quad (3.3)$$

The optimal paired solution (γ^*, O^*) for the above problem is obtained by a joint optimization procedure. We briefly present the two steps [3]:

[4] The orbit of a point in M refers to all possible points one can reach in M using the action of a group or a product of groups.

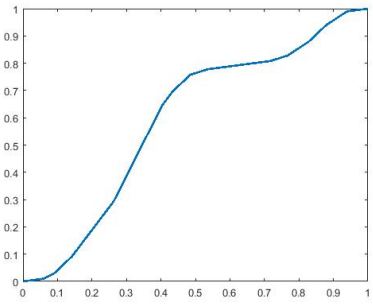


(a) *Artificial curves.*



(b) *Optimal Registration.*

(c) *Geodesic paths.*



(d) *optimal reparameterization.*

Figure 3.6: *Graphical illustration of the results obtained in the shape space between two artificial curves with different shape, scale and orientation.*

- **Optimal Rotation**

  Let's fix $\gamma \in \Gamma_1$, the optimization problem in equation (3.3) over $SO(3)$ is solved by the singular value decomposition USV^T of the matrix

  $$A = \int_0^1 q_1(t)(\sqrt{\dot{\gamma}(t)}q_2(\gamma(t)))^T dt$$

  that is $O^* = UV^T$. In case the determinant of A is negative, one needs to modify V by making its last column negative (of it current value) before multiplying to U to obtain O^*.

- **Optimal Registration**

  For a fixed O, optimization over Γ_1 is carried on with a DP algorithm, described in [3]. A gradient-based optimization may be considered as an alternative, with less accuracy but computationally less expensive.

In Figure 3.6 is shown an example of computing geodesics between artificial curves. It is clear the combination of bending and stretching betweem them using the estimated γ^*.

3.3.4 Joint Analysis with other Features

Under this setup, we are going to introduce new scenarios, each corresponding to a particular combination of physical properties of curves. For instance, when analyzing white matter fibers, features such as *scale* and *orientation* may be of interest.
We list now the different representative spaces, with their Riemannian structures and all the instruments for computing geodesic distances [3, 31].

1. **Shape and Orientation**

   Here again we rescale all the curves to a fixed length and we compute the SRVFs. The preshape space is still $\mathcal{C}_1$, while for the shape space we have to take care of the parameterization variability. Thus we introduce a new quotient space $\mathcal{S}_2 = \mathcal{C}_1/\Gamma_1$, which elements are the orbits given by $[q] = \{(q \circ \gamma)\sqrt{\dot{\gamma}} | q \in \mathcal{C}_1, \gamma \in \Gamma_1\}$. Note that we don't treat the orientation as a nuisance variable anymore.
   The optimization problems becomes:

   $$\gamma^* = \underset{\gamma \in \Gamma_1}{\operatorname{arginf}} \cos^{-1}(\langle q_1, \sqrt{\dot{\gamma}}(q_2 \circ \gamma)\rangle) = \underset{\gamma \in \Gamma_1}{\operatorname{arginf}} \left\| q_1 - \sqrt{\dot{\gamma}}(q_2 \circ \gamma) \right\|^2$$

   Then the geodesic path is given again by the great circle connecting $[q_1]$ and $[q_2]$ and the geodesic distance by $\theta = d_{\mathcal{S}_2}([q_1], [q_2]) = \cos^{-1}(\langle q_1, \sqrt{\dot{\gamma}}(q_2 \circ \gamma)\rangle)$.

2. **Shape and Scale**

   The size of the curves in the analysis is no longer relevant. This means that our SRVFs don't rely on a hypersphere, but they are still elements of the $\mathbb{L}^2$ space and the preshape space becomes $\mathcal{C}_2 = \left\{q : [0,1] \to \mathbb{L}^2([0,1], \mathbb{R}^3)\right\}$. The shape space is defined by $\mathcal{S}_3 = \mathcal{C}/(\Gamma_1 \times SO(3))$ and the orbits again by $[q] = closure\,\{O\sqrt{\dot{\gamma}}(q \circ \gamma) | \gamma \in \Gamma_1, O \in SO(3)\}$.
   The joint optimization is the same for the shape feature only, the difference here lies in the form of the geodesics. The geodesic path between elements of $\mathcal{S}_3$ is simply the straight line:

   $$\alpha(\tau)(t) = (1-\tau)q_1(t) + \tau q_2^*(t)$$

   and the geodesic distance is the norm of the difference, i.e $d_{\mathcal{S}_3}([q_1], [q_2]) = \left\| q_1 - O^* \sqrt{\dot{\gamma^*}}(q_2 \circ \gamma^*) \right\|$.

3. **Shape, Scale and Orientation**

   In case we are interested in comparing curves using their shapes, scales and orientations, the only variability we have to remove is the reparametrization. We form the quotient space as $\mathcal{S}_4 = \mathcal{C}_2/\Gamma_1$, which elements are the orbits $[q] = closure\,\{\sqrt{\dot{\gamma}}(q \circ \gamma) | \gamma \in \Gamma_1\}$.
   The optimization problem is $\gamma^* = \operatorname{arginf}_{\gamma \in \Gamma_1} \left\| q_1 - \sqrt{\dot{\gamma}}(q_2 \circ \gamma) \right\|^2$ and the geodesic distance between the registered curves is $d_{\mathcal{S}_4}([q_1], [q_2]) = \left\| q_1 - \sqrt{\dot{\gamma^*}}(q_2 \circ \gamma^*) \right\|$.
   The geodesic paths are again straight line $\alpha(\tau)(t) = (1-\tau)q_1(t) + \tau q_2^*(t)$.

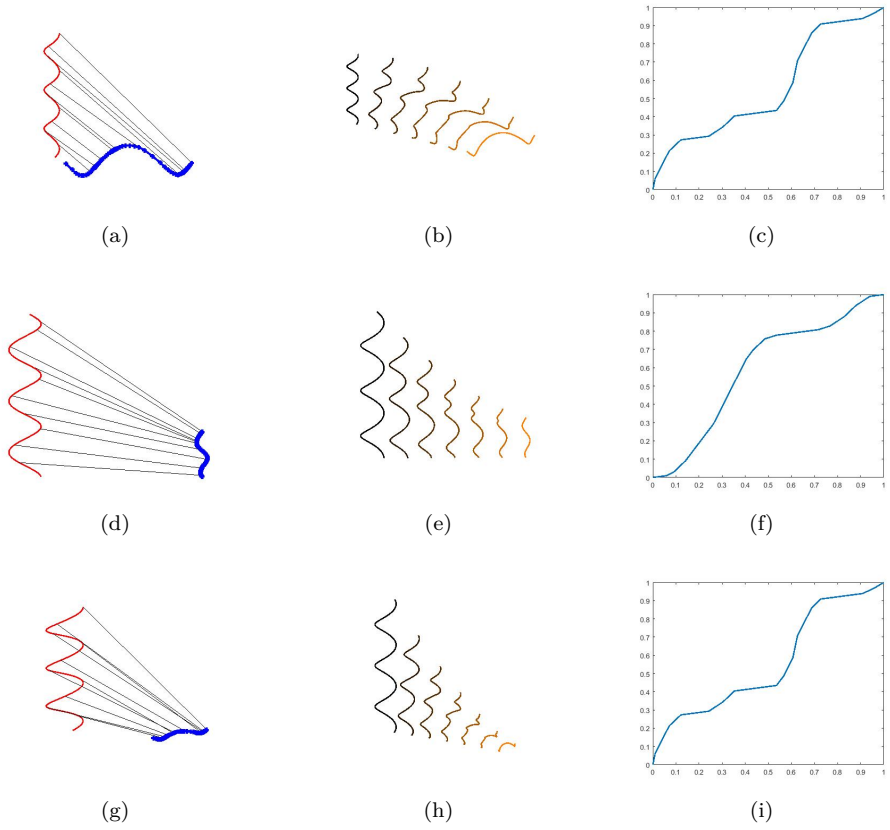


Figure 3.7: *The same two artificial curves of Figure 3.6 analyzed using different combination of features: (a),(b),(c) are the optimal registration, geodesic paths and reparameterization for Shape + Orientation; (d),(e),(f) for Shape + Scale; (g),(h),(i) for Shape + Orientation + Scale.*

Last consideration is about the inclusion of curve position in the analysis. In order to incorporate the position one can introduce a different representation, i.e. the **Square Root Function** (SRF). In our work we don't consider it, since it's not straightforward to recover the original curves from its SRFs. Thus the computation of geodesic path is not simple as in the SRVF case. While this drawback may limit the analysis, one can still obtain distances in a easy way. The shape shape is defined as $\mathcal{S}_{all} = \mathbb{L}^2([0,1],\mathbb{R}^3)/\Gamma_1$ and the distance is given by $d_{\mathcal{S}_{all}}([h_1],[h_2]) = \inf_{\gamma\in\Gamma_1} \|h_1 - \sqrt{\dot{\gamma}}(h_2 \circ \gamma)\|$, where the optimization is perfomed using again the DP algorithm.
In Figure 3.7 is shown a schematic view of the results obtained by different combination of features and in Table 3.1 we have summarized all the representative spaces and their characteristics.

Representative Spaces

Features	**Shape**	**Shape + Orientation**
Exp map	$exp_q(w) = \cos(\|\|w\|\|)q + \sin(\|\|w\|\|)\frac{w}{\|\|w\|\|}$	$exp_q(w) = \cos(\|\|w\|\|)q + \sin(\|\|w\|\|)\frac{w}{\|\|w\|\|}$
Exp^{-1} map	$exp_{q_1}^{-1}(q_2) = \frac{\theta}{\sin(\theta)}(q_2 - \cos(\theta)q_1)$	$exp_{q_1}^{-1}(q_2) = \frac{\theta}{\sin(\theta)}(q_2 - \cos(\theta)q_1)$
Distances	$d_{\mathcal{S}_1}([q_1],[q_2]) = \inf_{\gamma\in\Gamma_1, O\in SO(3)} \cos^{-1}(\langle q_1, O\sqrt{\dot{\gamma}}(q_2\circ\gamma)\rangle)$	$d_{\mathcal{S}_2}([q_1],[q_2]) = \inf_{\gamma\in\Gamma_1} \cos^{-1}(\langle q_1, \sqrt{\dot{\gamma}}(q_2\circ\gamma)\rangle)$
Geodesics	$\alpha(\tau) = \frac{1}{\sin(\theta)}(\sin(\theta(1-\tau))q_1 + \sin(\tau\theta)q_2^*)$	$\alpha(\tau) = \frac{1}{\sin(\theta)}(\sin(\theta(1-\tau))q_1 + \sin(\tau\theta)q_2^*)$
Features	**Shape + Scale**	**Shape + Orientation + Scale**
Exp map	$exp_q(w) = q + w$	$exp_q(w) = q + w$
Exp^{-1} map	$exp_{q_1}^{-1}(q_2) = q2 - q1$	$exp_{q_1}^{-1}(q_2) = q2 - q1$
Distances	$d_{\mathcal{S}_1}([q_1],[q_2]) = \inf_{\gamma\in\Gamma_1, O\in SO(3)} \|\|q_1 - O\sqrt{\dot{\gamma}}(q_2\circ\gamma)\|\|$	$d_{\mathcal{S}_2}([q_1],[q_2]) = \inf_{\gamma\in\Gamma_1} \|\|q_1 - \sqrt{\dot{\gamma}}(q_2\circ\gamma)\|\|$
Geodesics	$\alpha(\tau)(t) = (1-\tau)q_1(t) + \tau q_2^*(t)$	$\alpha(\tau)(t) = (1-\tau)q_1(t) + \tau q_2^*(t)$.

Table 3.1: Summary of the representative spaces and their intrinsic characteristics

Chapter 4

Applications in Statistical Modeling of Fibers

In the previous chapter we have defined a rigorous mathematical framework for the analysis of open curves in Riemannian spaces. This gives us the basic tools for computing summary statistics and for developing statistical models.
Now we are going to apply this approach to the modeling of fibers from the Left Corticospinal tract, that is computation of a representative fiber (the mean of the bundle), exploration of the variability inside the covariance structure, clustering to filter out outliers from the tractography and registration (pairwise or groupwise). The latter can also be perfomed in an augmented space including a scalar measure from the diffusion, for instance the FA values along the fibers.
This preliminary geometrical analysis paves the way to a challenging approach in functional data analysis, which will be faced in chapter 5.

4.1 Summary Statistics

In order to study white matter fibers, a reasonable choice is to consider shape in conjunction with scale and orientation. Therefore our analysis are mainly performed in the representative space $\mathcal{S}_4$, with some exceptions in other spaces to support better the ideas.
The very first step is to define and then estimate the first two central moments: the Karcher mean $\overline{\mu}_n$ and the covariance in the tangent space at the Karcher mean. In [3, 31], the Karcher mean is defined as:

$$\overline{\mu}_n = \underset{[q]\in\mathcal{S}_j}{\operatorname{argmin}} \sum_{i=1}^{n} d_{\mathcal{S}_j}([q],[q_i])^2, \quad j = 1, ..., 4 \tag{4.1}$$

A gradient-based approach is used to find a local minimum of the cost function. Here we present the general version of the algorithm, which can be adapted to the space j of interest:

Algorithm 4.1 (**Karcher Mean**). *Let $\{\beta_1, ..., \beta_n\}$ be a collection of curves and $\{q_1, ..., q_n\}$ its SRVFs. To minimize the cost function in (4.1), first let μ_0 be an initial estimate of the Karcher Mean (it seems natural to use the extrinsic mean, i.e $\mu_0 = \frac{1}{n}\sum_{i=1}^{n} q_i$ and project it if necessary). Set $j = 0$.*

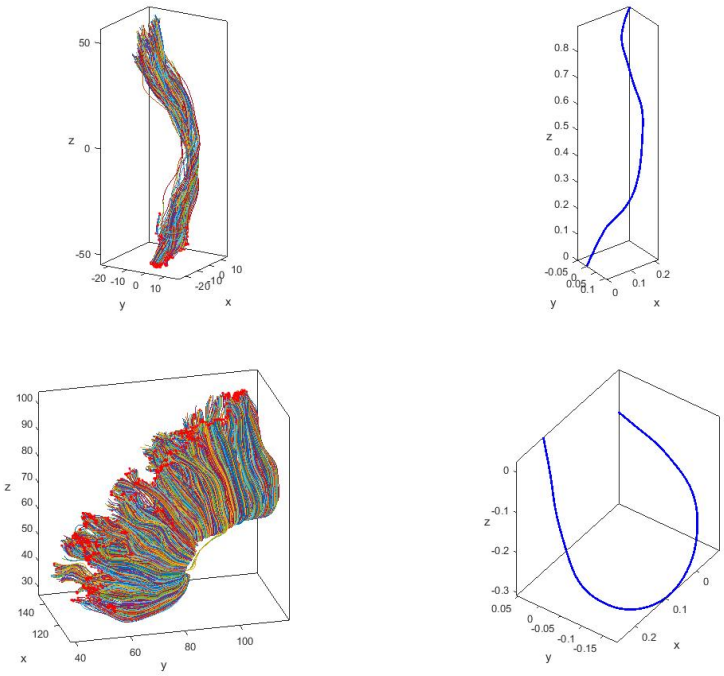


Figure 4.1: *Results of the application of the Karcher algorithm on the shape space $\mathcal{S}_1$. On the first row a Corticospinal tract, while in the second row a Genu from the Corpus Callosum. It's clear how the mean resembles the shape of the bundle.*

1. *For each $i = 1, .., n$ solve the specific optimization problem between μ_j and $[q_i]$. Once q_i^* is obtained, compute the tangent vector v_i*

$$v_i = exp_{\mu_j}^{-1}(q_i^*)$$

2. *Compute the average function $\overline{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$.*

3. *if $||\overline{v}||$ is small (good threshold may be 10^{-2}), then stop. Else, update μ_j in the direction $\overline{v}$*

$$\mu_{j+1} = exp_{\mu_j}(\epsilon\overline{v}), \quad \epsilon = 0.5$$

4. *set $j = j + 1$ and iterate.*

An example of computation of the Karcher mean for two different tracts is shown in Figure 4.1. Once we have the mean of our fiber bundle, we would like to explore the variability in the Karcher covariance structure. The computation is as follows [2, 31]:

1. Let v_i be the tangent vectors in $\mathbb{R}^3$ obtained in the last iteration of the above algorithm. With a slight abuse of notation, we concatenate the components in a tall vector $v_i \in \mathbb{R}^{3T}$, where T is the number of sampling points.

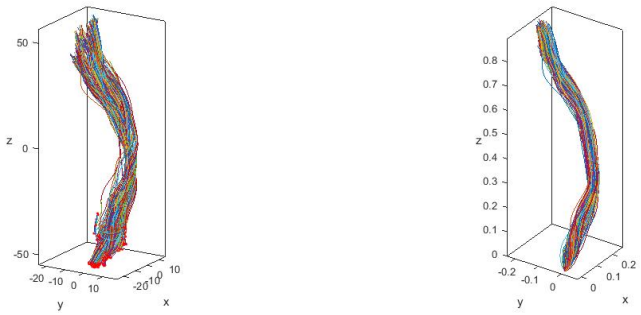


Figure 4.2: *Resampling strategy in the shape space $\mathcal{S}_1$. On the left the real tract, while on the right resampled fibers from a Wrapped Gaussian*

2. Compute the sample covariance matrix $K = \frac{1}{n-1}\sum_{i=1}^{n} v_i v_i^T$, which is a $3T \times 3T$ symmetric and non-negative matrix.

3. Let $K = U\Sigma U^T$ be its singular value decomposition. One in this way can reach an efficient basis using traditional PCA. The submatrix formed by the first r columns of U spans the principal subspace of the observed data.

Following this procedure we can visualize variations in curves along principal paths and also perform a sampling method from a Wrapped Gaussian Distribution. We illustrate this approach considering only the shape space $\mathcal{S}_1$, which simplifies the analysis. Given the SVD decomposition of the Karcher covariance, we can introduce a multivariate Gaussian model for the long vector $v \in R^{3T}$, i.e. we define $v = \sum_{i=1}^{n} z_i\sqrt{\Sigma_{ii}}U_i$, where $z_i \sim N(0,1)$ iid. One can then rearrange the vector v in its components and project it back to the shape space using the exponential map to obtain a random curve. This is a useful technique in bootstrapping statistics. In Figure 4.2 we can see how this technique works well, with satisfactory results in preserving the original shape of the bundle.
In a similar manner we can also explore the variation along the principal paths by computing straight lines along dominant directions. We do this for the same data used for resampling. In Figure 4.3 are shown curves rendered along the path $exp_{\overline{q}}(\tau v)$ for τ from -1 to $+1$, where $v = \sqrt{\Sigma_{ii}}U_i$, while in Figure 4.4 a compact visualization on the $3D$ space is displayed.

4.2 Hierarchical Clustering

An important problem deriving from tractography estimation is the detection of outliers, which may affect the analysis and the reliability of the results. In our case, we would like to make use of the tools described in the previous chapters. Since we have defined different distances between curves under different combination of features, we can use these distances for clustering. Here we don't aim to cluster different bundles in the brain, but rather to clean out a precise bundle from outliers.
We experiment this idea considering the representative space $\mathcal{S}_4$, since the combination of shape, orientation and scale features gives better discrimination than shape information alone. It should be pointed out that the clustering in this case is solely

Figure 4.3: *Variation along the first mode from $\tau = -1$ to $\tau = 1$, with the middle curve representing the mean $\overline{q}$*

Figure 4.4: *The first and second eigenmodes of the shape, represented as deviation from the Karcher Mean.*

data-driven and it makes sense, whereas we want to filter out outliers. We exploit the dendrogram clustering program in MATLAB with an average linkage (the mean distance between elements of each cluster).
In Figure 4.5, the upper-left corner shows an original tract with a group of fibers which are most probably outliers due to the different shape, while the upper-right is the result from clustering. We can iterate this procedure to reach a desired result, as shown in the bottom row.

4.3 Registration

The problem of joint registration and shape analysis can be studied either as pairwise or groupwise. In the first case, one solves for which point on one curve matches which point on the other. Since the two curves are parameterized, the registration is controlled by parameterization: for any $t \in I = [0, 1]$, the points $\beta_1(t)$ and $\beta_2(t)$ are

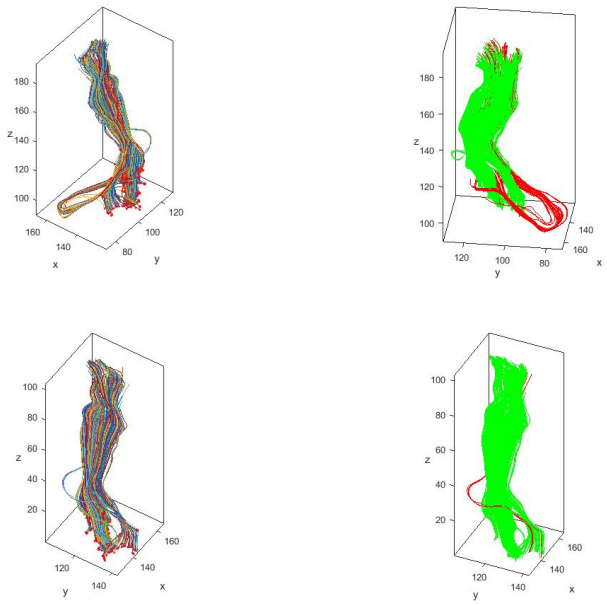


Figure 4.5: *Results of the clustering algorithm (two iterative steps) with the combination of Shape, Orientation and Scale. In red we can see the fibers which have been filtered out.*

registered. As we have seen in the optimization problem, the change in registration is accomplished by a warping function.
The groupwise registration problem, instead, is that given multiple curves $\{\beta_1, ..., \beta_2\}$ we look for corresponding reparameterization functions $\{h_1, ..., h_2\}$ such that $\{f_i(h_i(t))\}$ are registered for all $t \in I$. The idea is to compute a mean shape (i.e., the Karcher Mean) and to align under a proper metric the individual curves to this mean shape [3]. The innovative aspect is the inclusion of a texture information in the problem. In case there is any additional information associated with curves, other than the usual features, it can be used to help improve the comparisons and registration between curves. This motivates us to introduce a space of augmented curves, where we would like to perform the process of matching, deformation and comparison to be based on the representative space and the auxiliary information [36]. All this while keeping appropriate constraints and invariances.
Let $\beta_s(t) : [0, 1] \to \mathbb{R}^3$ be the usual open curve and $\beta_t(t) : [0, 1] \to \mathbb{R}^k$ the auxiliary function. We can combine these two components to form a curve:

$$\beta(t) = \begin{bmatrix} \beta_s(t) \\ b\beta_t(t) \end{bmatrix} \in \mathbb{R}^{3+k}$$

Here $b > 0$ may be a parameter introduced to control the influence of the auxiliary function, relative to the shape function. In our case the auxiliary function is the FA values along the fibers (so $k = 1$), which can be used to obtain a better alignment within a bundle. For the resulting augmented curve, we define the SRVF using the usual formula and the *augmented preshape*:

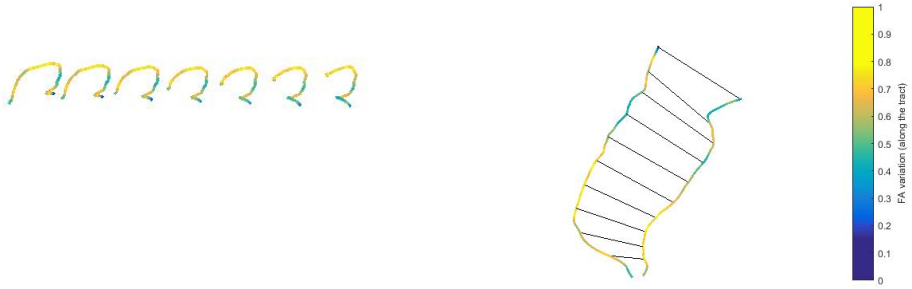


Figure 4.6: *Geodesic paths and correspondances in the augmented space (FA as the auxiliary information) for two random fibers.*

$$\mathcal{C}_{aug} = \left\{ q : [0,1] \to \mathbb{R}^4 | \int_0^1 |q(t)|^2 dt = 1 \right\} \tag{4.2}$$

As usual, we endow it with a Riemannian metric. For a point $q \in \mathcal{C}_{aug}$ and any two tangents $w_1, w_2 \in T_q(\mathcal{C}_{aug})$, we define the inner product:

$$\langle w_1, w_2 \rangle = \int_0^1 \langle w_1, w_2 \rangle dt$$

where the inner product in the integrand is the standard Euclidean one. We have now to treat the desired invariances in the analysis.

- **Scaling**: The global scaling has been accounted in (4.2). If a different scaling is needed for the texture function, it can be controlled by the constant b.
- **Rotation**: we may be interested to remove the orientation from the representation, while maintaining the texture component the same. Thus we define:

$$R = \begin{bmatrix} SO(3) & 0 \\ 0 & 1 \end{bmatrix}$$

  in such a way we only modify the shape component.

- **Translation**: The definition of q itself provides the removal of position information in the analysis. In most of the cases, one needs to bring back the translation of the texture information β_t in the representation. This can be easily done including a constant vector, which may be for instance the mean value of the texture function along the curve:

$$\overline{\beta}_t(t) = \frac{\int_0^1 \beta_t(t) dt}{\int_0^1 dt} = \int_0^1 \beta_t(t) dt \in \mathbb{R}^k$$

We recover the scalar function after registration substracting the new mean and adding the mean saved before.

- **Reparametrization**: we seek to use both features and auxiliary information in performing registration. As usual, we first define an equivalence class under reparametrization and we optimize to find γ^*.

In [27], for instance, this approach is applied to the study of the shapes of carotid arteries, which are analyzed as elastic curves. In our framework, the registration strategy points to align fibers and scalar function along them, exploiting both coordinates and texture information together. In figure 4.6 we can see an example of pairwise registration with the FA feature included as a texture information, which has been performed in the $\mathcal{S}_4$ space. The geodesic paths for the auxiliary information are simply given considering a linear interpolation of the average texture $\overline{\beta}(\tau) = (1-\tau)\beta_1 + \tau\beta_2^*$. Instead, in Figure 4.7 we can clearly see the effect of the groupwise registration in terms of better alignment in phase and amplitude of the FA functions. This procedure can be iterated until a desired result is obtained.

4.4 Summary

In this chapter we have discussed the main applications of shape analysis in statistical modeling of fibers. In certain kind of diseases, the computation of distances between them may help to figure out possible damaged areas in the brain. The possibility to study the shapes of fibers and to associate them with functionality between brain regions is a prominent research question.
In our case we want to pull together shape and functional approach in order to investigate scalar measures from diffusion along fibers. With tools from shape analysis we aim to model a specific tract (i.e., Corticospinal) for each patient, computing the Karcher Mean and registering the fibers to it including the FA information. Then we move to the analysis of these FA functions trying to figure out a priori known lesions. This is the starting goal of the next chapter.

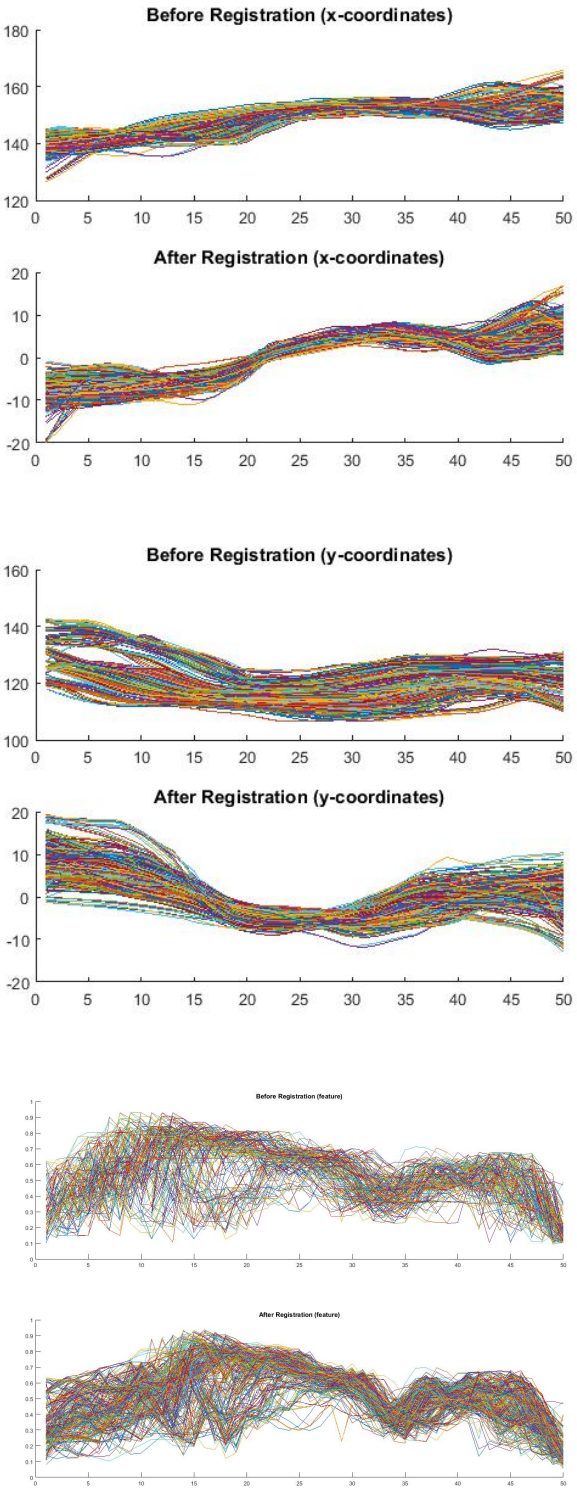


Figure 4.7: *Comparison between the x-coordinates (above), y-coordinates (in the middle) and the feature (below) before and after groupwise registration to the population Mean. This procedure can be iterated until a desired result is obtained.*

Chapter 5

A New Perspective with Functional Data Analysis

5.1 Goals and Proposals

We have already highlighted how Multiple Sclerosis can affect the human brain. Lesions on the brain may vary from patient to patient and are not predictable. One possible solution is to analyze tract profiles in order to localize abnormalities in subjects that can be manifested by significative changes in scalar measures from diffusion. It is recognized that FA index is sensitive to pathological damages and in active lesions its values may decrease according to the severity of the disease [7].
By this point, we should also be aware that a basic truth doesn't exist. Inconsistencies may exist with regard to FA results and different factors can affect our analysis. Nevertheless, the study of diffusion indeces along fibers represents a challenging approach to the problem. In Figure 5.1 we can see an example of the FA profile along the left Corticospinal tract, while in Figure 5.2 is shown the localization of a lesion on it for the same patient.
A question naturally arises: *is it possible to detect these lesions working on the FA functions?* Several previous works have faced similar questions. In [25] diffusion indeces along fiber tracts are measured as function of geodesic distance from specific anatomical landmarks, while in [16] a tract-based analysis is introduced. Furthermore, in [4] a functional approach is considered: diffusion properties are modeled as multivariate functions of arc length and hypothesis testing is performed in the projection on the PCA space.
In this thesis we would like to perform a tract-based analysis in a functional framework, exploiting two different approaches:

- **Interval-Wise Testing Procedure**: this approach relies on statistical inference in a functional framework. In [26], Pini and Vantini propose a non-parametric inferential procedure able to detect portions of the domain where a null hypothesis is rejected in a functional sense. The most common situation is a test on mean differences between two groups.

- **Functional Linear Discriminant Analysis**: here, instead, we want to exploit the ability of this technique to produce a discriminant function able to identify areas of discrimination between classes, i.e. between patients and controls in our

case. In [10], James and Hastie propose an extension of the linear discriminant approach to a functional framework, accounting also for curves irregularly sampled along the domain.

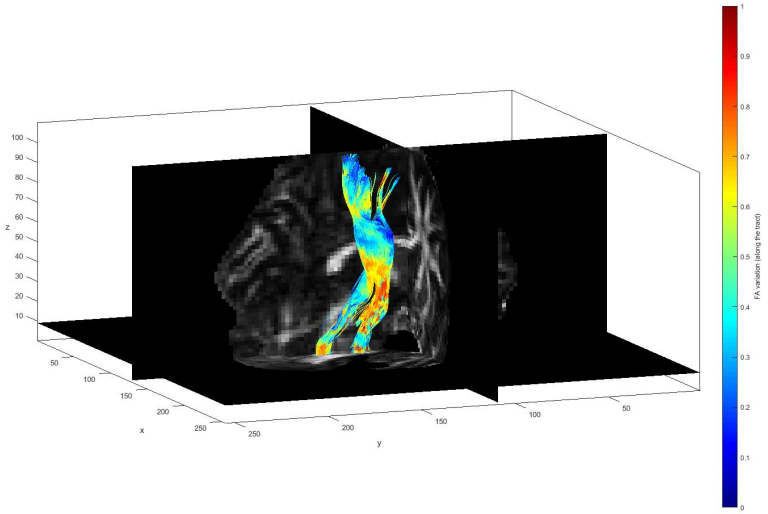


Figure 5.1: *FA profile along the CST plotted on a MRI volume*

Figure 5.2: *Localization of a lesion on the CST for a patient affected by Multiple Sclerosis*

5.2 Preparing the Data

In Chapter 2 we have seen how the raw data from diffusion is processed and how the extraction of the fiber bundle of interest is computed with the software WMQL. Once we have obtained the tract for each patient and control under analysis, we read it in Matlab using open source codes described in [13]. These tools let us to visualize the fibers in a 3D fashion, to resample them with a cubic spline interpolation (we have choosen $T = 50$ sampling points [1]) and to extract scalar measures along them given a MRI volume, as displayed in Figure 5.1. To read the volume, for Linux machine and mac OS, one can run FSL commands from Matlab (function *read avw*), while in Windows there is a tool for Nifti images that makes almost the same work.
The uniform number of vertices along the fibers, as proposed in [13], accounts for inter-streamline and inter-subject scaling and it facilitates the analysis.
Afterwards, as the first step in our Shape analysis, we try to filter out outliers from tractography with the hierarchical clustering algoritm. Once the bundle is cleaned, we model it in the $\mathcal{S}_4$ space, computing the Karcher mean and the scalar mean for each individual. In this way we obtain a vector $\mu(t) \in \mathbb{R}^4$, where the components are the x, y, z, coordinates and the FA value. To align the fibers we compute a population mean, as shown in Figure 5.3, and we apply the registration procedure highlighted in section 4.3. This results in phase alignment with respect to a common mean for all the individuals, both patients and controls.
We now save the aligned functions and we move to R for performing functional data analysis. In the case of IWT, a preliminary smoothing procedure is not necessary and we can directly apply the procedure on the raw functions. Instead, for the FLDA, we looked for the number of spline basis that best suits our data.

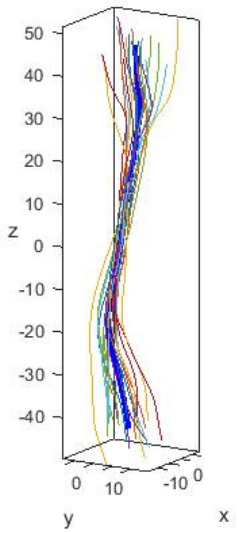


Figure 5.3: *The population mean highlighted in blue for patient and control group.*

[1] The choice of the number of sampling points is related with the size of the image, in our case $128 \times 128 \times 55$

5.3 Interval-Wise Testing Procedure

We first give a brief summary of the methodology in order to understand the inferential approach in a FDA framework. Here again the infinite-dimensionality of the problem poses new challenges: evaluation of point wise p-values is meaningless in a functional space and the multiplicity correction would involve a family of infinite tests. We describe the procedure in a version adapted to our case.
Suppose we have a set of $\mathbb{L}^2$ scalar functions over a domain $T = (a, b) \subset \mathbb{R}$ and we want to perform a global unilateral hypothesis test in a two-sample framework of the following form

$$H_0 : \mu_1 = \mu_2 + C \;\; vs \;\; H_1 : \mu_1 > \mu_2 + C \tag{5.1}$$

where $\mu_j, j = 1, 2$ are the mean functions for the two groups, C is a constant and the equality has to be intended in $\mathbb{L}^2$ sense. In case of rejection of the null hypothesis H_0, one may be interested on the portion of the domain presenting a significant mean difference. To achieve this goal we introduce restrictions (i.e. open intervals $\mathcal{I} = (t_1, t_2) \subset T$, with $a_1 \leq t_1 < t_2 \leq a_2$) on the domain and we formulate

$$H_0^{\mathcal{I}} : \mu_1^{\mathcal{I}} = \mu_2^{\mathcal{I}} + C \;\; vs \;\; H_1^{\mathcal{I}} : \mu_1^{\mathcal{I}} > \mu_2^{\mathcal{I}} + C$$

This implies that $H_0 = \bigcap_{\mathcal{I} \subset T} H_0^{\mathcal{I}}$. Often assumptions of functional normality are not realistic and a non-parametric permutation procedure is more suitable. For the general case of a bilateral test, in [26] a statistic based on the $\mathbb{L}^2$ distance between the two means is used:

$$T^{\mathcal{I}}(\xi_{11}, ..., \xi_{1n_1}, \xi_{21}, ..., \xi_{2n_2}) = \frac{1}{|\mathcal{I}|} \int_{\mathcal{I}} (\overline{\xi}_1(t) - \overline{\xi}_2(t))^2 dt$$

where ξ_{ji} are the functional observations and $\overline{\xi}_j(t) = \frac{1}{n_j} \sum_{j=1}^{n_j} \xi_{ji}$. In our case, we modify the test statistic above to account for the unilaterality of the test. A possible choice may be to take $T^{\mathcal{I}}$ without the square and to consider just the positive part of it, as follows

$$T^{\mathcal{I}}(\xi_{11}, ..., \xi_{1n_1}, \xi_{21}, ..., \xi_{2n_2}) = \begin{cases} \frac{1}{|\mathcal{I}|} \int_{\mathcal{I}} (\overline{\xi}_1(t) - \overline{\xi}_2(t)) dt, & \text{if } T^{\mathcal{I}} > \; 0 \\ 0, & \text{otherwise} \end{cases} \tag{5.2}$$

Finally, a permutation test for (5.2) can be reached by evaluating the test statistic over all possible permutations of our data.
At this point, we need to introduce the concept of p-value function in this framework. One of the major contribution in [26] is the definition of an unadjusted and adjusted p-value with well-suited properties. Let $p^{\mathcal{I}}$ denote the restricted p-value of the functional test described above, then we have

- **unadjusted p-value**: it controls the **point-wise error** (given any point where H_0 is not violated, the probability of wrongly selecting it as significant is controlled) and is defined as

  $$p(t) = \limsup_{\mathcal{I} \to t} p^{\mathcal{I}}$$

  where with the notation $\mathcal{I} \to t$ we intend that both the extremes of the interval $\mathcal{I}$ converges to t.

- **adjusted p-value**: it controls the **interval-wise error** rate (control of the probability of wrongly select an interval of the domain where H_0 is not violated) and is defined as

$$\widetilde{p}(t) = \sup_{\mathcal{I} \ni t} p^{\mathcal{I}}$$

These definitions bring important theoretical properties. Here we list them, for more details on its proof see [26]. We can state that, even though the pointwise evaluation is meaningless in a infinite dimensional framework such as $\mathbb{L}^2$ space, the boundedness of the p-values $p^{\mathcal{I}}$ guarantee that they are well-defined $\forall t \in T$. Furthermore, if we consider an open limited interval $T \in \mathbb{R}$, we have that $\mathbb{L}^2(T) \subset \mathbb{L}^1(T)$ and thus the integral mean value theorem guarantees that

$$p(t) = \lim_{\mathcal{I} \to t} p^{\mathcal{I}}$$

and it coincides almost everywhere with the p-value of the permutation test based on the statistic (5.2). Moreover, in the special case of data embedded in $\mathbb{L}^2(T) \cap C^0(T)$, the identity above holds $\forall t \in T$.
From a practical point of view, the p-value is evaluated as the proportion of the corresponding test statistics that exceed the statistics on the original data set and by thresholding it at level α we select the intervals of the domain presenting significant differences in the mean. In detail, to control the point-wise error rate we select the points $t \in T$ such that $p(t) \leq \alpha$. Instead, if one is interested in controlling the interval-wise error rate, we select the points $t \in T$ such that $\widetilde{p}(t) \leq \alpha$.

5.3.1 Detection of a Clinically Significant Threshold

The constant C in (5.1) plays the role of the detector. The idea is to find out a suitable value for this constant able to highlight portions of the domain with significant differences for a patient with a lesion on the CST.
We perform a sort of heuristic binary classification exploiting the Interval-Wise Testing procedure. In detail, we consider $n_1 = 1$ patient and $n_2 = 10$ controls and the pairwise functional test in (5.1) assumes the form of a two class-prediction problem as follows:

- $\frac{n_2(n_2-1)}{2}$ IWTs between the controls, which represents our True positive condition (P). Ideally, we aim to validate the null hypothesis for the controls, i.e. no differences in the FA means for each subject.

- n_2 IWTs between the patient and the controls, which represents our True negative condition (N). In this case we expect significant evidence against the null hypothesis if a lesion is present on the CST.

Specifically, for each threshold, we first apply the IWT between controls in a pairwise manner (*Control1* vs *Control2*, *Control1* vs *Control3* and so on) and we consider the result as **False Negative** if we detect differences in the domain, while **True Positive** if not. Afterwards we test the patient against each control (*Patient* vs *Control1*, *Patient* vs *Control2* and so on) and we consider the result as **True Negative** if we find differences, while **False Positive** if not. We consider the patient with a big lesion on the CST depicted in Figure 5.2 and we carry out the above procedure. For instance, let's take as the first threshold $C = 0$, i.e. no mean difference for the

		PREDICTED CONDITION	
		Pred. positive	**pred. negative**
TRUE CONDITION	**Cond. positive** (P)	True positive (TP)	False negative (FN)
	Cond. negative (N)	False positive (FP)	True negative (TN)

Table 5.1: Confusion matrix for a two class-prediction problem.

FA profile in the hypothesis test (5.1). In this case we find significant differences also between most of the controls, since clinically speaking the FA profile may vary from subject to subject due to different factors. The aim is to increase step by step the value of C and to find the first threshold which best discriminate between the patient with lesion and the group of controls. We then build iteratively the confusion matrix as shown in Table (5.1) and we select the best threshold exploiting the ROC curve.
The latter is a graphical plot that illustrates the performance of a binary classifier system when its discrimination threshold is varied. In our case, the threshold is simply the mean difference we test in the functional hypothesis procedure. The curve is created by plotting the **true positive rate** (TPR or sensitivity) on the y-axis and the **false positive rate** (FPR or specivity) on the x-axis at various threshold settings. Following the confusion matrix in table (5.1), the TPR and FPR are computed for each threshold as follows:

$$TPR = \frac{TP}{P}, \quad FPR = \frac{FP}{N}$$

How do we select the best value for the mean difference? The best prediction would yield a point in the upper left corner of the ROC space. In practice , the latter is divided by a diagonal line, where points above it represent good classification results, while points below it represent poor results. We select as the best threshold the point closer to $(0, 1)$ coordinates, which in our case is given by $C = 0.09$, as depicted in Figure 5.4. To validate it, we compare the patient under analysis and a patient without lesion against the same control. The selected threshold highlights significant dissimilarities in the middle-upper part for the patient with lesion, while any difference is detected for the other patient, as shown in Figure 5.5.

5.4 Functional Linear Discriminant Analysis

We start by presenting the model on which the FLDA procedure is based. We don't go deep into the details, since it's beyond our scope. For further information on the derivation of this method the reader can refer to [10].
The main idea is to model each functional observations using a spline basis multiplied by a q-dimensional coefficient vector. Let $g_{ij}(t)$ be the true value at time t for the jth curve from the ith class. Obviously we have a finite number of time points for each curve, i.e. $t_{ij1}, ...t_{ijn_{ij}}$ (n depends on ij since in the general case we may have irregularly sampled curves, but for us is fixed). Then we have

$$Y_{ij} = g_{ij} + \epsilon_{ij}, \;\; i = 1, ..., K, \;\; j = 1, ..., m_i$$

where K is the number of classes and m_i the number of curves in the ith class. The errors ϵ_{ij} are assumed to have zero mean , constant variance σ^2 and be uncorrelated between each other and g_{ij}. Now we choose the cubic spline functions and we define

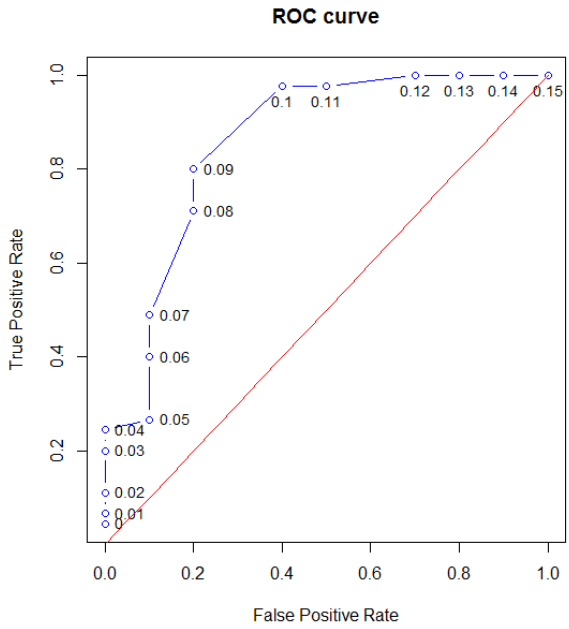


Figure 5.4: *The ROC curve for the detection of the optimal threshold.*

$$g(t) = S(t)^T \eta$$

where $S(t)$ is a spline basis in matrix form $n \times q$ and η is a q-dimensional vector of spline coefficients. This leads to a more restricted model, which in index form is as follows

$$Y_{ij} = S_{ij}\eta_{ij} + \epsilon_{ij}, \quad i = 1, ..., K, \quad j = 1, ..., m_i$$

Now the problem of modeling the observations reduces to one of modeling the coefficients η_{ij}. A natural extension of the LDA suggests to use a Gaussian distribution. Finally, we obtain the following formulation

$$\begin{aligned} Y_{ij} = S_{ij}(\mu_i + \gamma_{ij}) + \epsilon_{ij}, \quad i = 1, ..., K, \quad j = 1, ..., m_i \\ \epsilon_{ij} \sim N(0, \sigma^2 I), \quad \gamma_{ij} \sim N(0, \Gamma) \end{aligned} \tag{5.3}$$

One of the tasks of LDA is to trasform or project the variables into a lower dimensional subspace and then classifying in this subspace. In this setting, Anderson (1951) and Hastie and Tibshirani (1996) outline an alternative procedure that accounts for this. They give raise to the final FLDA model defining the mean in (5.3) as $\mu_i = \lambda_0 + \Lambda\alpha_i$, where λ_0 and α_i are respectively q and h-dimensional vectors and Λ is a $q \times h$ matrix, $h < \min(q, K)$. Certain kinds of constraints that accounts for the normalization of linear discriminants need to be imposed. The final formulation given in [10] is

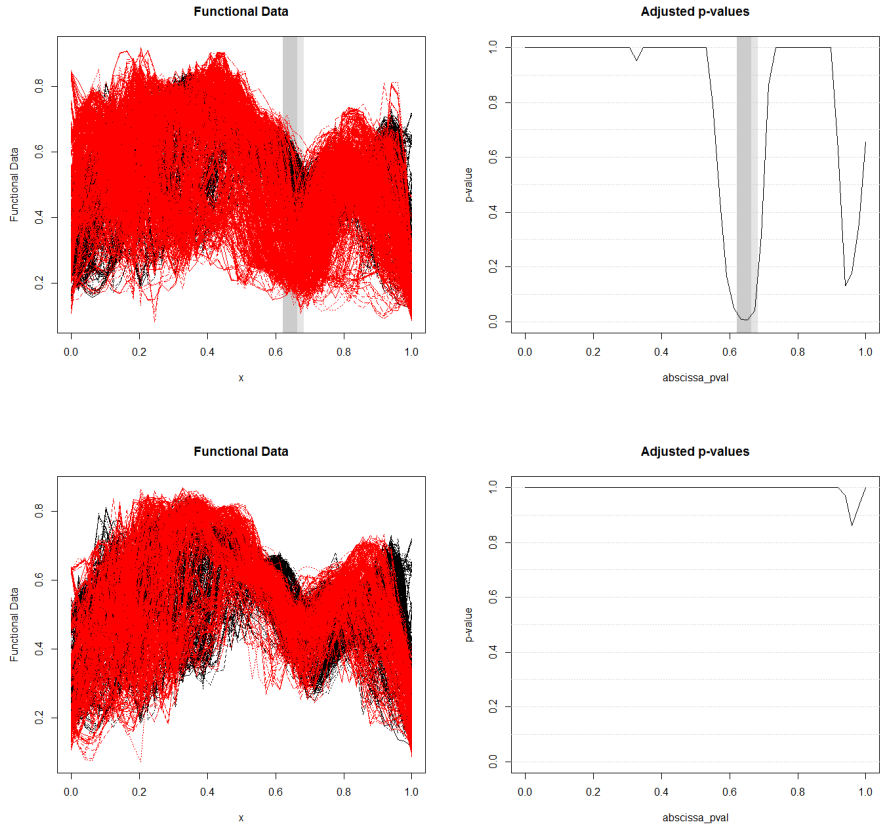


Figure 5.5: *Results from IWT with the best threshold for a patient with a damaged CST (above) and a patient without lesions (below) against the same control.*

$$Y_{ij} = S_{ij}(\lambda_0 + \Lambda\alpha_i + \gamma_{ij}) + \epsilon_{ij}, \quad i = 1, ..., K, \quad j = 1, ..., m_i$$
$$\epsilon_{ij} \sim N(0, \sigma^2 I), \quad \gamma_{ij} \sim N(0, \Gamma)$$
$$\Lambda^T S^T \Sigma^{-1} S\Lambda = I, \quad \sum_i \alpha_i = 0$$

where $\Sigma = \sigma^2 I + S\Gamma S^T$ and S is the basis matrix evaluated over a fine grid of points.
Fitting the model involves the estimation of the unknown parameters $\lambda_0, \Lambda, \alpha_i, \Gamma, \sigma^2$ and a natural approach to perform this is to maximize the likelihood of Y_{ij}. Since this represents a difficult non-convex optimization problem, in [10] is proposed instead to maximize jointly the likelihood of Y_{ij} and γ_{ij}, which semplifies the analysis.
We also underline that, as in most of the functional analysis approaches, the choice of q (i.e. the dimension of the spline basis) affects the procedure. Thus a smoothing analysis is needed before applying the FLDA model.

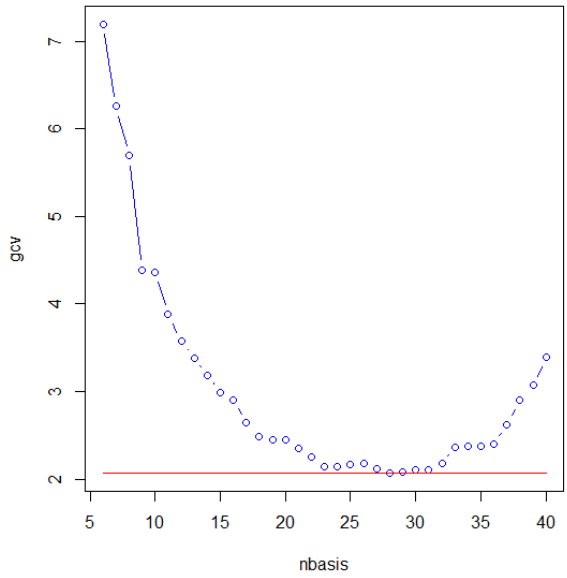


Figure 5.6: *Plot of the generalized croos-validation against the number of q.*

5.4.1 Applications

Now we are going to see how this model can be applied to the analysis of FA functions. The main idea is to discriminate areas of dissimilarities represented by known lesions in a specific tract. We apply FLDA in order to further validate the result obtained in the previous section, i.e. we consider again the patient with lesion against a control. The first step is the choice of a reasonable number q for the spline basis and the *generalized cross-validation* (GCV) measure is designed to locate it. In our analysis, a suitable range for q is $[25, 30]$. For instance, for the patient with lesion, $q = 28$ reaches the minimun in the GCV as shown in Figure 5.6.
The second step is the fitting of the model, which involves the estimation of the unknown parameters. Figure 5.7 shows the fitted FA functions for the two subjects, while in Figure 5.8 we can see the mean and the two centroids, which are obtained as follows

$$\mu_i(t) = S(t)^T(\lambda_0 + \Lambda\alpha_i), \quad i = 1, 2$$

We can now proceed to the estimation of the discriminant function. In a standard two-class LDA the latter is defined as

$$(\mu_1 - \mu_2)^T\Sigma^{-1}$$

where Σ^{-1} is the within group covariance matrix. In a functional framework, considering that $\mu_i = S(\lambda_0 + \Lambda\alpha_i)$, the analogue is

$$(S\Lambda\alpha_1 - S\Lambda\alpha_2)^T\Sigma^{-1} = (\alpha_1 - \alpha_2)^T\Lambda^T S^T\Sigma^{-1} \tag{5.4}$$

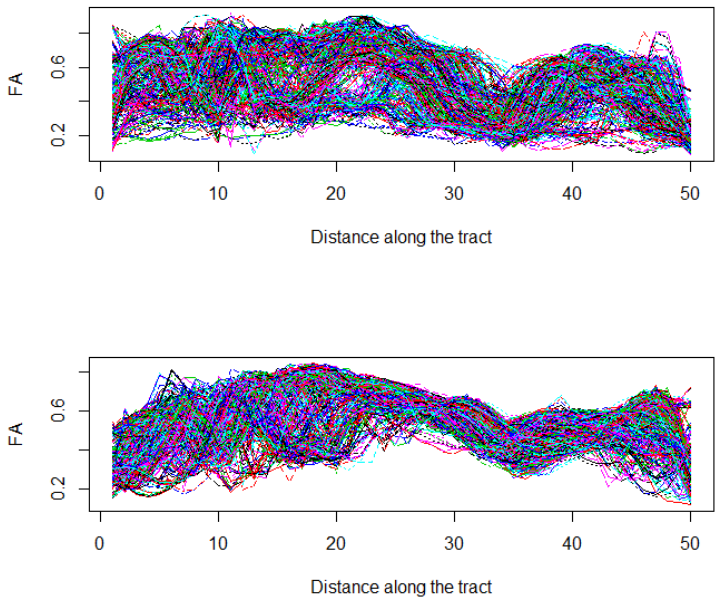


Figure 5.7: *FA functions for the patient (above) and the control (below) with a suitable number of spline basis.*

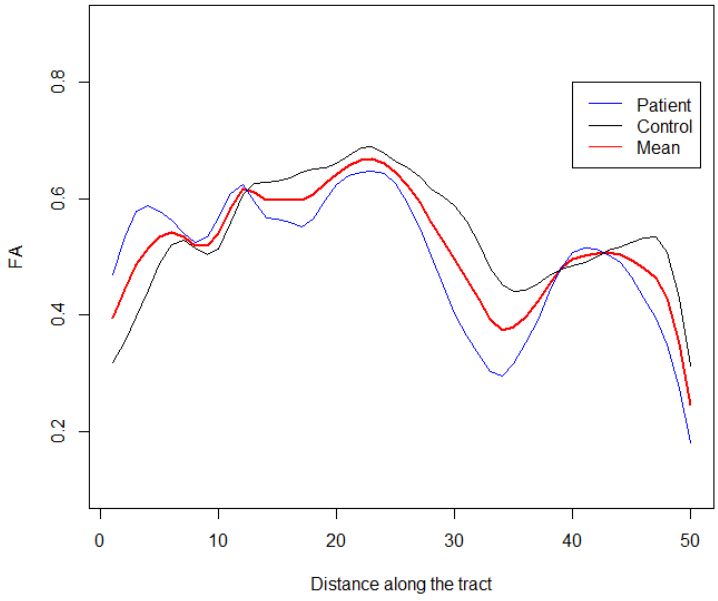


Figure 5.8: *Mean function and centroids.*

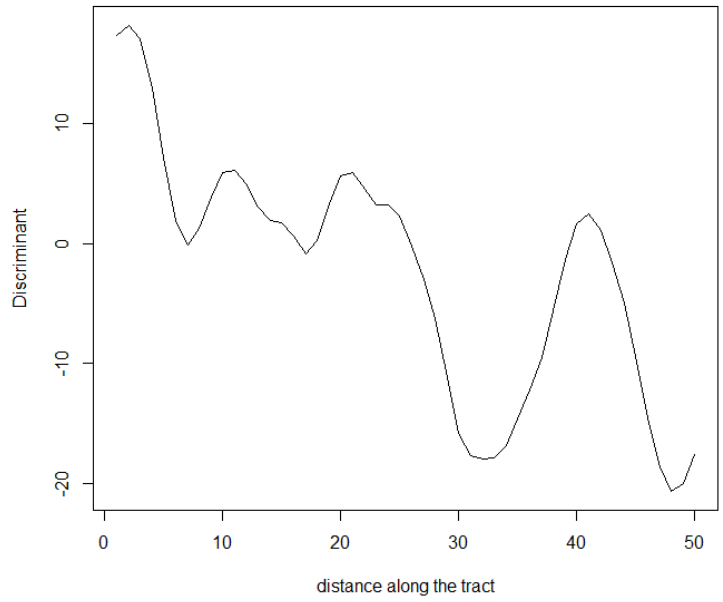


Figure 5.9: *Discriminant function for the patient with lesion and the control.*

	Patient	Control
Patient	565	40
Control	17	511

Table 5.2: Missclassification table for the FLDA. We have $n_1 = 605$ FA functions for the patient and $n_2 = 528$ for the control.

Figure 5.9 shows the discriminant function produced by Eqn (5.4) using the state of the subject (patient or control) as the class variable. We can see that most of the discrimination appears in the middle upper-part with a strong negative peak followed by a positive one. This fact may indicate the strong difference in FA profile in proximity of the lesion, as detected with the IWT procedure.
The ability of FLDA to perform classification is as important as its ability to explain discrimination between different classes. While classification is not our primary goal, we test the prediction strength of the model on the data in order to assess the reliability of the results.
For a two class-situation, such as our case, the classification rule is based on the negativity or positivity of the discriminant scores for each observations. Figure 5.10 shows the discriminant scores plotted on the x-axis and the mean FA value along each function on the y-axis . The two dotted vertical lines either side of zero give the fitted values for the α_i's. The percentage of the missclassified functions is 0.0503%, computed through the missclassification table in 5.2.

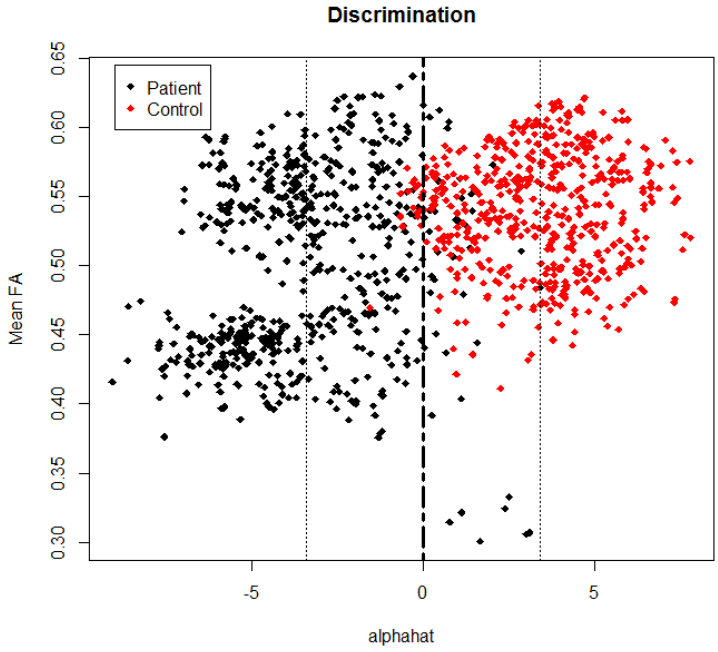


Figure 5.10: *Plot of the discriminant scores. The separation of the two classes is net.*

Chapter 6

Conclusions and Perspectives

In this work we focused mainly in the statistical modeling of fibers with tools from shape analysis. As described in chapter 3, one can combine different features depending on the kind of problems. This approach can find various applications, such as improving the results from tractography and studying the connectivity between different areas of the brain.
In our case, shape analysis allowed us to model the geometry of the left corticospinal tract. We have computed summary statistics, such as the Karcher mean and the covariance structure. The latter can be explored to study the variability of the tract in terms of principal components (PCA). The ability to compute distances between fibers is of great importance, since they can be used for clustering and for evaluating dissimilarities/similarities among different subjects. Regarding registration, we showed two different strategies: pairwise and groupwise. Specifically, we extracted FA along fibers in order to investigate possible abnormalities in patients with lesions on the CST and we registered them to a population mean. In this way we tried to pull together shape and functional approach in an efficient way.
Once we have registered all the FA functions for each subject, we faced the problem of detecting changes in the FA profile in presence of a lesion known a priori in a Multiple Sclerosis research study. The methods proposed aimed to find a significant threshold in the mean difference of FA values using statistical inference in a functional framework. Subsequently, with a linear discrimination approach, we tried to confirm the results obtained.
Possible issues may derive from the basic steps of the analysis, where semplicistic methods in fiber tractography are used. For this reason, scalar measures can be affected by multiple fibers in the same voxel or even worse by some parts which are missing.
The ideas presented in chapter 5 are thought as possible starting points for a more detailed and deep analysis in a functional framework. The possibility to investigate directly diffusion properties on a specific tract is an open and interesting challenge in the medical community. Most of the previous statistical methods considered an average approach, accounting for a single mean function. Here we proposed a more efficient tract-based approach.
We do not claim to assess the general validity of our proposals. Indeed, for a matter of time, we have showed the results on just a patient. Possible future directions may be in generalizing and further investigating this context. In particular, future developments may regard:

- optimize the tractography step, with more sophisticated model that may result

in more robustness in the extraction of the tracts.

- improve the inference procedure for the detection of a significative threshold, considering a greater number of patients.
- moving from a subject-to-subject to a subject-to-group analysis in the functional linear discriminant analysis. Clinically speaking, the first may give poor results.

Neverthless, the link between Shape and Functional data analysis in the quantitative study of white matter fibers in clinical diseases represents a novel method and could adress these problems in a stimulant perspective.

Bibliography

[1] Anuj Srivastava et Al., *Shape Analysis of Elastic Curves in Euclidean Spaces*, IEEE Trans Pattern Anal Mach Intell. 2011 Jul; 33(7):1415-28. Epub 2010 Oct 14.

[2] Anuj Srivastava et Al., *Statistical Shape Analysis: Clustering, Learning, and Testing*, IEEE Trans Pattern Anal Mach Intell., Vol. 27, No. 4, April 2005.

[3] Anuj Srivastava and Eric P. Klassen, *Functional and Shape Data Analysis*, Springer Series in Statistics, 1st edition, 2016.

[4] Casey B. Goodlett et Al., *Group analysis of DTI fiber tract statistics with application to neurodevelopment*, NeuroImage 45 (2009) S133-S142.

[5] Christopher G. Small, *Statistics of shape*, John Wiley & Sons, Volume 3 September/October 2011.

[6] Demian Wassermann et Al., *The white matter query language: a novel approach for describing human white matter anatomy*, Brain Struct Funct (2016), Springer-Verlag Berlin Heidelberg.

[7] Emilia Sbardella et Al., *DTI Measurements in Multiple Sclerosis: Evaluation of Brain Damage and Clinical Implications*, Hindawi Publishing Corporation, Multiple Sclerosis International Volume 2013.

[8] Fernanda Tovar-Moll et Al., *Diffuse and Focal Corticospinal Tract Disease and its Impact on Patient Disability in Multiple Sclerosis*, J Neuroimaging. 2015 March; 25(2): 200-206.

[9] G. C. DeLuca et Al., *Axonal loss in multiple sclerosis: a pathological survey of the corticospinal and sensory tracts*, Brain (2004), 127, 1009-1018.

[10] Gareth M. James and Trevor J. Hastie, *Functional Linear Discriminant Analysis for Irregularly Sampled Curves*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2001.

[11] Heidi Johansen-Berg and Timothy E.J. Behrens, *Diffusion MRI: From quantitative measurement to in vivo neuroanatomy*, 1st edition, 2009.

[12] Ivan I. Maximov et Al., *Robust tensor estimation in diffusion tensor imaging*, Journal of Magnetic Resonance 213 (2011) 136-144.

[13] John B. Colby et Al., *Along-tract statistics allow for enhanced tractography analysis*, NeuroImage 59 (2012) 3227-3242.

[14] John M. Lee, *Introduction to Smooth Manifolds*, Second edition, Springer: Graduate Texts in Mathematics, 2012.

[15] Karl-Olov Löblad, *MR Imaging in Multiple Sclerosis: Review and Recommendations for Current Practice*, AJNR Am J Neuroradiol 31:983-89, Jun-Jul 2010.

[16] Lauren J. O'Donnell, Carl-Fredrik Westin and Alexandra J. Golby, *Tract-based morphometry for white matter group analysis*, NeuroImage 45 (2009) 832-844.

[17] Marc Niethammer et Al., *On Diffusion Tensor Estimation*, Conf Proc IEEE Eng Med Biol Soc. 2006; 1:2622-2625.

[18] Marvin M. Goldenberg, *Multiple Sclerosis Review*, P&T. 2012 Mar; 37(3): 175-184.

[19] Meena Mani et Al., *A Comprehensive Riemannian Framework for the Analysis of White Matter Fiber Tracts*, 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Rotterdam, 2010, pp. 1101-1104.

[20] Mikkel B. Stegmann and David Delgado Gomez, *A Brief Introduction to Statistical Shape Analysis*,

[21] P. Mukherjee et Al., *Diffusion Tensor MR Imaging and Fiber Tractography: Theoretic Underpinnings*, AJNR Am J Neuroradiol 29:632-41, 2008.

[22] Pablo G.P. Nucifora et Al., *Diffusion-Tensor MR Imaging and Tractography: Exploring Brain Microstructure and Connectivity*, Radiology: Volume 245: Number 2-November 2007.

[23] Patric Hagmann et Al., *Understanding Diffusion MR Imaging Techniques: From Scalar Diffusion-weighted Imaging to Diffusion Tensor Imaging and Beyond*, RG Volume 26, October 2006.

[24] Peter J. Basser and Derek K. Jones, *Diffusion-Tensor MRI: theory, experimental design and data analysis - a technical review*, NMR Biomed. 2002;15:456-467.

[25] Pierre Fillard et Al., *Quantitative Analysis of White Matter Fiber Properties along Geodesic Paths*, R.E. Ellis and T.M. Peters (Eds.): MICCAI 2003, LNCS 2879, pp. 16-23, 2003.

[26] Pini Alessia and Vantini Simone, *The interval testing procedure: A general framework for inference in functional data analysis*, Biometrics 2016 Sep; 72(3):835-45.

[27] Qian Xie, Sebastian Kurtek and Anuj Srivastava, *Analysis of AneuRisk65 data: Elastic shape registration of curves*, Electronic Journal of Statistics, Vol. 8 (2014) 1920-1929.

[28] R. Wang et Al., *Diffusion Toolkit: A Software Package for Diffusion Imaging Data Processing and Tractography*, Proc. Intl. Soc. Mag. Reson. Med. 15 (2007).

[29] Saad Jbabdi and Heidi Johansen-Berg, *Tractography - where do we go from here?*, Brain Connect. 2011 ; 1(3): 169-183.

[30] Sebastian Kurtek et Al., *Elastic Geodesic Paths in Shape Space of Parameterized Surfaces*, IEEE transactions on Pattern Analysis and Machine Intelligence (2012 vol. 34)

[31] Sebastian Kurtek et Al., *Statistical Modeling of Curves Using Shapes and Related Features*, Journal of the American Statistical Association, September 2012.

[32] Serge Lang, *Differential and Riemaniann Manifolds*, Third edition, Springer: Graduate Texts in Mathematics, New York, 1995.

[33] Shantanu H. Joshi et al., *A Novel Representation for Riemannian Analysis of Elastic Curves in* $\mathbb{R}^n$, Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2007 July 16; 2007(17-22 June 2007): 1-7.

[34] Susumu Mori and Peter C. M. van Zijl, *Fiber tracking: principles and strategies - a technical review*, NMR IN BIOMEDICINE NMR Biomed. 2002; 15:468-480.

[35] Susumu Mori and Jiangyang Zhang, *Principles of Diffusion Tensor Imaging and Its Applications to Basic Neuroscience Research*, Neuron 51, 527-539, September 7, 2006, Elsevier Inc.

[36] Wei Liu, Anuj Srivastava and Eric Klassen, *Joint Shape and Texture Analysis of Objects Boundaries in Images Using A Riemannian Approach*, 42nd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2008, pp. 833-837.