



HAL
open science

Automated In Vivo Dissection of White Matter Structures from Diffusion Magnetic Resonance Imaging

Demian Wassermann

► **To cite this version:**

Demian Wassermann. Automated In Vivo Dissection of White Matter Structures from Diffusion Magnetic Resonance Imaging. Medical Imaging. Université Nice Sophia Antipolis, 2010. English. NNT: . tel-00827022

HAL Id: tel-00827022

<https://theses.hal.science/tel-00827022>

Submitted on 28 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PhD THESIS

prepared at

INRIA Sophia Antipolis

and presented at the

University of Nice-Sophia Antipolis

Graduate School of Information and Communication Sciences

*A dissertation submitted in partial satisfaction
of the requirements for the degree of*

DOCTOR OF SCIENCE

Specialized in Signal and Image Processing

Automated *In Vivo* Dissection of White Matter Structures from Diffusion Magnetic Resonance Imaging

Demian Wassermann

Advisor	Dr. Rachid Deriche	INRIA Sophia-Antipolis, France
Reviewers	Dr. Cyril Poupon	CEA NeuroSpin, France
	Dr. Carl-Fredrik Westin	Harvard University, USA
Examiners	Dr. Peter Basser	National Institute of Health, Bethesda, USA
	Dr. Habib Benali	INSERM / Pitié-Salpêtrière, France
	Dr. Ragini Verma	University of Pennsylvania, USA

THÈSE

préparée à l'

INRIA Sophia Antipolis

et présentée à l'

Université de Nice-Sophia Antipolis

École Doctorale des Sciences et Technologies de la
Communication et de l'Information

pour obtenir le titre de

DOCTEUR EN SCIENCES

Spécialité : Traitement du Signal et des Images

Dissection Automatique *In Vivo* des Structures de la Matière Blanche à partir de l'Imagerie par Résonance Magnétique de Diffusion

Demian Wassermann

Directeur de thèse	Dr. Rachid Deriche	INRIA Sophia-Antipolis, France
Rapporteurs	Dr. Cyril Poupon Dr. Carl-Fredrik Westin	CEA NeuroSpin, France Harvard University, USA
Examineurs	Dr. Peter Basser Dr. Habib Benali Dr. Ragini Verma	National Institute of Health, Bethesda, USA INSERM / Pitié-Salpêtrière, France Université of Pennsylvania, USA

A Marcos

ACKNOWLEDGEMENTS

I can inscribe these last years as some of the best of my life, a tale that started from the moment I stepped into a plane in Buenos Aires in order to perform research internship in France. This is, without doubt, due to many people I encountered in the ever-travelling, even drifting, life I lead during my PhD. Even if in this small text I am only talking about some of them, I have fond memories of uncountable encounters with marvelous persons which happened during these past three, almost four, years. For all of this, I am grateful.

First of all I would like to thank my supervisor, Rachid Deriche. Not only for his academic advising but for his moral and human guidance. The first mail I received from him when I arrived to Sophia-Antipolis for a four-month internship read "Welcome...home !!!" was, unexpectedly, a sign of things to come. Even though the circumstances didn't picture a standard PhD programme scenario, he took me into his team and opened the world's doors for me. He has not only provided me with valuable academic tools to continue my career but also made efforts towards positioning me in the academic world. While working with him, I could appreciate his academic and human skills and learn from them. A good example of these skills is in the way in which he transmits knowledge, which ranges from discussing mathematical equations to telling fables when the occasion calls for it. During these years, he has become a good friend with whom I share coffees, BBQs and his Czech beers. Last but not least, I admire the way in which he invests the same amount of energy in the academic leadership of his team as in turning it into a second family for its members. This is further shown by the way in which he sees off his students, as children leaving the nest towards their "maturity".

I also want to acknowledge Cyril Poupon and Carl-Fredrik Westin for their constructive reviews of my thesis and Peter Basser, Habib Benali and Ragini Verma for accepting to on this thesis jury, I am waiting for your comments and feedback anxiously.

I am thankful to my teammates in the Odyssee (currently Athena) research group. Particularly, Christophe Lenglet for welcoming me; Max Descoteaux for mentoring me, sharing our life abroad in morning coffees and having lots of fun during academic travels; John Touboul, Romain Veltz, François Grimbert, and Adrien Whorer for guiding my maths; Alexandre Gramfort for coping with my talking compulsion in the office and Sylvain Vallaghe for the ski lessons and Ma-Jo Escobar (and family) and Sandrine Chemla for the shared moments. All of them made not only a good team but dear friends during my PhD. Lastly, thanks to the next generation, Auro Ghosh and Manu Caruyer and to the "experienced ones" Théo Papadopoulo and Maureen Clerc for helpful discussions and interchanges.

A warm thanks to all the people who kindly hosted me into their labs. They showed me how there are several angles to the same problems and in some cases they became close friends: CF Westin, Steve Pieper, Raul San Jose-Estepar, Gordon

Kindlmann and Lauren O'Donnell at Boston; Ragini Verma, Luke Bloy and Stathis Kanterakis at Philadelphia; Stéphane Lehèricy, Habib Benali, Cristine Delmaire, Julien Cohen-Hadad, and Arnaud Messé at Paris; Alfred Anwander and Thomas Knoschë at Leipzig and Mariano Sigman and Bruno Wicker at Buenos Aires. Moreover, I am grateful to Guillermo Marshall and Esteban Mocksos and the all of the complex systems lab of the school of exact sciences at the University of Buenos Aires for supporting me in Argentina and providing me with the means to start my academic career.

Additionally, this thesis would have not been possible without the help and support and laughs of many new and old friends the last few years. At Buenos Aires, Ariela, Caro, Diana and Mariana far away, yet present. The ones dancing with me Karine, Costanza, Julie, Heloïse and Caro, who are a spread around the world now. The french guys, Tom Veracauteren who surprised me by reading through all of my thesis and Stanley Durrleman who showed me that it is great to be generous even if you are competing. The antibois Alex, Sap, Flo and the Eganaude volleyball pack.

A special thanks to Rebecca Wolpin for exhaustively proofreading and editing most of the chapters of this thesis and greatly contributing to the quality of this manuscript. I am also grateful to Emmanuel Caruyer and Sylvain Merlet for the French translation of the introduction and conclusion.

Finally, I am deeply grateful to my parents Débora and Alfredo and my sister Eliana. Without their help, support, and trust through the happy and the tough times I would have never gotten here. I am sincerely indebted with you three.

Thanks, to all of you.

Dem

Contents

Acknowledgements	5
i Introduction	11
Context	11
Organization of This Thesis	12
ii Résumé	17
Contexte	17
Plan du manuscrit et contributions	19
I Background	23
1 Brain History	25
1.1 Ancient History: The seat of the mind	25
1.2 Middle Ages: Small rooms	32
1.3 Renaissance: The human being is the centre	33
1.4 Modern Age: Gross tractography	36
1.5 Nineteenth Century: Chemical Tractography	41
1.6 Twentieth Century: Imaging-Based Tractography	43
1.7 Automated <i>In Vivo</i> Dissection of the White Matter	44
1.8 Summary	46
2 Brain Anatomy and Diffusion MRI	47
2.1 Human Brain and Neural Tissue	47
2.2 The Diffusion Process	53
2.3 Nuclear Magnetic Resonance	56
2.4 Diffusion-Weighted MRI of Living Tissue	60
2.5 Tractography	72
2.6 Summary	84
3 Clustering	87
3.1 What is Clustering?	87
3.2 Clustering Algorithms	92

3.3 Summary	103
II White Matter Structure Identification Through Manifold Learning	105
4 Manifold Learning	117
4.1 Feature matrix-based techniques	117
4.2 Proximity matrix-based techniques	120
4.3 Summary	132
5 White Matter Structure Identification	133
5.1 Spectral Embedding and Clustering	133
5.2 Materials	136
5.3 Experiments and results	137
5.4 Discussion	140
5.5 Summary and Conclusion	144
III Gaussian Processes Modelling of the White Matter Fibres	147
Our Contribution	154
6 Gaussian Processes	155
6.1 Non-linear Regression with a Bayesian flavour	155
6.2 Non-Stationary Priors on Function Regression	161
6.3 Function inner product space	164
6.4 Summary	168
7 Gaussian Process Framework for White Matter Bundles	169
7.1 Intuition	169
7.2 Gaussian Process Framework	172
7.3 Clustering Algorithm	178
7.4 Results	180
7.5 Discussion	182
7.6 Conclusion	190
8 Tractography-Based Spinal-Cord Straightening	191
8.1 Introduction	191
8.2 Material and methods	192
8.3 Results	196
8.4 Discussion and Conclusions	197

9	Statistical Analyses of the White Matter in Schizophrenia	199
9.1	Introduction	199
9.2	Statistical Analysis Algorithms	200
9.3	Data	203
9.4	Results	205
9.5	Discussion	205
9.6	Summary	206
	Appendix	224
A	Publications of the Author	227
	Journal Publications	227
	Conferences and Workshops with Proceedings	227
	Conference Abstracts	228
B	Useful Mathematical properties	229
	B.1 Properties of Gaussian distributions	229
	B.2 Linear Algebra	229
	Bibliography	231

CHAPTER I

INTRODUCTION

CONTEXT

How does the brain work? This has been a recurrent question through history.

The first documented causal relationship between speech, motor tasks and the brain dates from 3,000 BCE [Wilkins: 1964]. However, the role of the brain remained a subject of debate until Galen of Pergamon (127- 199), based on several predecessors, convinced the western civilization that it was the centre of sensations and the intellect. With the advent of the European Renaissance, the use of images to document and study human anatomy began to play a role in the study of the brain through the hands of Leonardo Da Vinci (1452-1519) and Andreas Vesalius (1514-1564) [Linden: 2002, Gross: 1998b]. In particular, Vesalius [1563] distinguished for the first time the soft tissue on the cortex of the brain from the hard white substance in its interior [Singer: 1956]. Marcello Malpighi (1628-1694) studied this white substance and, in using a primitive microscope, discovered that it was composed of fibres. Extending this work, Nicholaus Steno (1638-1686), argued that in order to discover the purpose of these fibres, their pathways should be traced and studied. However, it wasn't until years later that Felix Vicq D'Azyr differentiated association from commissural fibres and proposed that the multiple connections had a role in keeping the normal functioning of the brain in the case of "inconveniences". Gall and Spurzheim [1810] took this a step further, they stated that the cortex had areas with specialized functions and the white matter was the "wiring" connecting these areas, a theory that revolutionised neuroscience. This was further proved by Broca [1861] who showed that damage in a precise area of the cortex caused speech impediments. The paradigm of cortical specialization along with the newly included neuron theory gave place to the importance of the human brain's white matter in the functioning brain [Schmahmann and Pandya: 2007a]. This led Dejerine and Dejerine-Klumpke [1895] and Wernicke to produce detailed descriptions of the white matter in their atlases and to link changes in the white matter structure structure with neurological disorders. During the twentieth century, theories about brain function began to afford a more prominent place to white matter leading to, and influenced by, the

development of several methods to trace the tracts of the white matter were developed [Schmahmann and Pandya: 2007a]. However, it was not until the introduction of Diffusion Magnetic Resonance Imaging (DMRI) [Taylor and Bushell: 1985, Websey et al.: 1984a,b, Le Bihan and Breton: 1985, Le Bihan et al.: 1986] and tractography that the analysis of white matter could be done *in vivo* and non invasively [Basser et al.: 2000, Mori et al.: 2005, Catani et al.: 2002].

The great success of DMRI comes from its capability to accurately describe the geometry of the underlying microstructure [Johansen-Berg and Behrens: 2009]. DMRI captures the average diffusion of water molecules, which probes the structure of the biological tissue at scales much smaller than the imaging resolution. The diffusion of water molecules is Brownian under normal unhindered conditions, but in fibrous structures such as white matter, water molecules tend to diffuse along fibres. Due to this physical phenomenon, DMRI is able to obtain information about the neural architecture *in vivo*. Furthermore, the introduction of DMRI-based tractography, a technique to trace the white matter fibre bundles *in vivo*, opened new perspectives in studying the anatomy of the white matter and relating it to neurological disorders [Ciccarelli et al.: 2008].

This thesis builds on research DMRI and tractography [Basser et al.: 2000, Mori et al.: 2005, Catani et al.: 2002, Behrens et al.: 2003, Lenglet et al.: 2006b, Descoteaux et al.: 2007a]. DMRI is not only able to model and analyse the structure of the white matter within a voxel, but that it is also a powerful tool to analyse the anatomy of the brain and its connectivity. Moreover, it has been recently proved useful in order to analyse the differences on the white matter and its consequences in brain function [Ciccarelli et al.: 2008, Kubicki et al.: 2007].

The starting point and motivation of this thesis is the *in vivo* dissection of the white matter, performed by Mori et al. [2005] and Catani et al. [2002]. This procedure isolates the white matter tracts that play a role in a particular function or disorder of the brain so they can be analysed looking for differential characteristics. Manually performing this dissection requires a great knowledge of brain anatomy and several hours of work. Hence, the development of a technique able to automatically perform the identification of white matter structures is of utmost importance.

ORGANIZATION OF THIS THESIS

This thesis is organized in three parts. The first part describes the white matter cerebral history and anatomy, the principles of DMRI and the main algorithmic techniques required to understand this thesis. Then, the second part describes the first methodological contributions of this thesis: the identification of white matter structures through clustering QBall imaging voxels. The third part describes the contributions at the heart of our methodological development to automatically perform the identifications of white matter structures from tractography: the proposal of a Gaussian process mathematical framework for white matter fibre bundles, its application

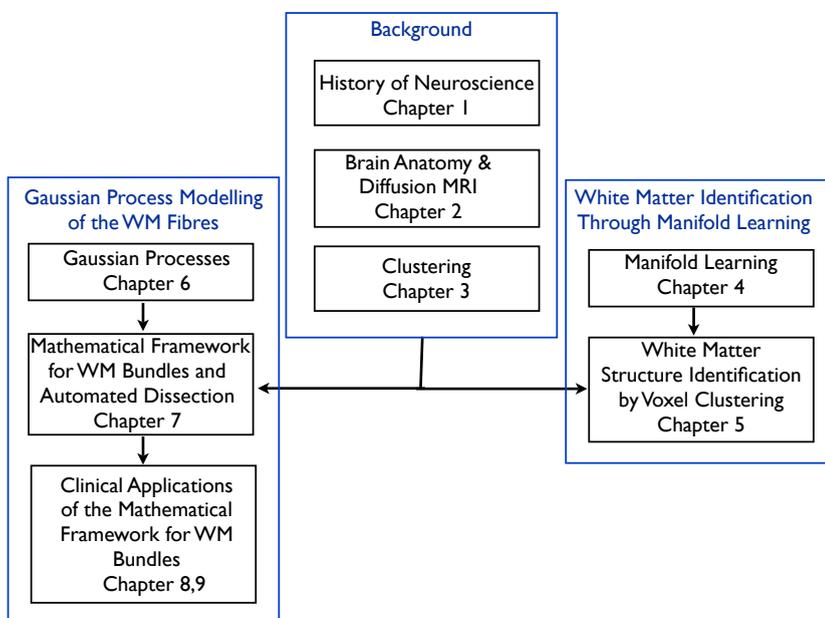


Figure i.1: Sketch of the chapters of this thesis.

to automated *in vivo* dissection, spinal cord straightening and population differences in schizophrenia. The organization of this thesis shown in figure i.1.

Part I: Background

The background part describes the white matter cerebral anatomy, the principles of DMRI and tractography and finally a review of clustering techniques.

Chapter 1 briefly reviews the history of neuroscience. It starts from the first account of the relationship between the brain and intellectual and motor tasks and ends in the present day, where the current development of magnetic resonance imaging (MRI) renders possible to analyse the brain *in vivo* and non-invasively.

Chapter 2 answers the following question: what are we looking for in the human brain white matter? In particular, what is the organization of the white matter? How is the brain connected? What are the large fibre bundles in the brain? How can we analyse it through DMRI? This chapter covers the basic cerebral anatomy of white matter and its imaging through DMRI. First, we review brain organization and we highlight the three major white matter tract groups and present some of its most representative tracts. Then, we introduce the physical process of diffusion and how it is related to the structure of living tissue followed by a brief review of the history of nuclear magnetic resonance imaging, from its beginning until the development of

DMRI. Finally, the chapter describes MRI-based tractography, a tool capable of reconstructing white matter fibres bundles, and some of its major clinical applications.

Chapter 3 answers the following question: Given a set of elements, how can we get insight on them and divide them into groups? The main goal of this thesis is to identify white matter structures; to this end, techniques designed to group elements into clusters are fundamental. This chapter builds a taxonomy of clustering algorithms dividing them into families and describes the major representatives of each family.

Part II: White Matter Structure Identification Through Manifold Learning

Part II deals with the identification of white matter structures directly from the images obtained by DMRI. In order to do this, we present an algorithm which combines techniques from two areas: manifold learning and clustering. The first technique is about how to automatically find a good representation for a set of objects, in our case voxels from DMRI images and the second technique deals with the clustering of this voxels.

Chapter 4 answers the question: given a set of elements, is it possible to automatically reduce the complexity of their representation by inferring a suitable space for them? The representation of the water diffusion profile within a voxel is complex. Hence, algorithms which aim to perform statistical analysis of DMRI images must deal with the complexity of this representation. This chapter reviews state-of-the-art manifold learning techniques. These techniques take a set of elements as an input and infer a representation of their features which is simpler to handle than the original. Manifold learning techniques are fundamental for the contribution presented as they enable us to automatically simplify the information of the DMRI images while keeping its most important features.

Chapter 5 addresses the question: is it possible to segment a DMRI image into several white matter bundles automatically? This chapter presents a new clustering algorithm to segment high-angular resolution diffusion imaging (HARDI) data. More precisely, in this chapter, we compare the use of DT and QBI images and show what can be gained by using HARDI data. We also discuss how the combination of manifold learning techniques and clustering is able to provide a tool suited to simultaneously identify several white matter bundles. First, we use Markovian relaxation in order to seamlessly integrate spatial and diffusion information. Then, the diffusion maps manifold learning technique is used in order to infer a suitable representation of the image voxels and improve the posterior clustering results. Finally, a clustering algorithm is applied to the output of the previous step. Results

on synthetic and real human brain images are presented in order to demonstrate the advantages of this algorithm.

Part III: Gaussian Process Modelling of White Matter Fibres

Part III deals with the main contribution of this thesis: the proposal of a mathematical framework for white matter fibre bundles. Here we describe the needed mathematical background, the framework itself and its application to automatic *in vivo* dissection of white matter.

Chapter 6 answers the following question: What is an appropriate mathematical tool to represent white matter fibre bundles? This chapter covers the mathematical concepts needed to understand the major contribution of this thesis. In particular, how Gaussian processes are an appropriate mathematical tool to represent smooth functions, curves and bundles of curves in space. Finally, it covers how to use this framework to quantify the similarity between smooth functions and how to define an inner product space of these functions such that it is possible and straight forward to perform statistical analysis of these functions.

Chapter 7 addresses the following: How can the Gaussian Process framework be applied to the representation of white matter fibre bundles? Is it possible to use the inner product space provided by this representation in combination with anatomical information in order to perform automatic *in vivo* dissection of the white matter? In this chapter we show that Gaussian processes are a fit model for white matter fibre tracts. Moreover we take advantage of the inner product space of white matter fibre bundles represented as Gaussian processes: using this space, we develop a clustering algorithm which is capable of incorporating anatomical information. Finally, we apply this algorithm to a database of 68 subjects and we identify successfully the major white matter structures of the brain.

Chapter 8 illustrates other uses of the Gaussian process-based modelling of white matter fibres. Within this chapter, we provide a sound formulation to select the most representative fibre of a bundle. Then, we use this formulation in order to develop a fully automated algorithm to straighten spinal cord images using DMRI-based tractography. Straightening of spinal cord images constitutes an important application. The raw, non-straightened visualization of spinal cord images may hamper diagnosis because the curvature of the spinal cord in the antero-posterior direction makes it difficult to obtain its full coronal or sagittal picture. Moreover, this curvature is subject-dependent and therefore group studies are hard to achieve at the spinal level. Consequently, a method to straighten the spinal cord such that individual images can be compared an eventually pooled constitutes an important application of DMRI-based tractography.

Chapter 9 illustrates the uses of our Gaussian-process framework in order to perform statistical analyses among populations. we combine the clustering results obtained from the algorithm presented in chapter 7 with a dataset composed of healthy controls and schizophrenic subjects to show group differences in the white matter fibre bundles of the brain. By expressing state-of-the-art statistical analysis tools in terms of our framework, we are able to study group differences between schizophrenic and healthy subjects. These studies are carried on in the tracts as a whole and then in a skeletonized representation in order to show differences in precise areas.

CHAPTER II

RÉSUMÉ

CONTEXTE

Au cours de l'histoire, l'homme s'est posé la question du fonctionnement du cerveau de façon récurrente. La première trace écrite d'une mise en évidence de la relation entre le cerveau d'une part, et le langage et les tâches motrices d'autre part remonte à l'an 3000 avant J.-C. [Wilkins: 1964]. Cependant, le rôle joué par le cerveau est longtemps resté sujet à caution, jusqu'à ce que Galien de Pargame (127-199), se basant sur les travaux de certains prédécesseurs, convainque la civilisation occidentale que le cerveau constitue le centre des perceptions sensorielles et de l'intelligence. À l'arrivée de la Renaissance en Europe, l'utilisation de documents illustrés pour l'anatomie a commencé à s'imposer dans l'étude du cerveau, sous l'impulsion de Léonard de Vinci (1452-1519) et d'André Vésale (1514-1564) [Linden: 2002, Gross: 1998b]. En particulier, Vésale [1563] a été le premier à distinguer les tissus mous constituant le cortex cérébral, de la matière blanche plus rigide qui constitue la partie interne du cerveau [Singer: 1956]. Marcello Malpighi (1628-1694) a étudié cette substance blanche et a découvert à l'aide d'un microscope primitif qu'elle était composée de fibres. Dans la continuité de ces travaux, Nicolas Sténon (1638-1686) a émis l'idée que pour comprendre la fonction de ces fibres, il serait nécessaire de tracer et d'étudier leurs trajectoires. Ce n'est que des années plus tard que Félix Vicq d'Azyr a fait pour la première fois la distinction entre fibres commissurales et fibres d'association, et il a émis l'hypothèse que ces connexions multiples jouaient un rôle dans le maintien des fonctions cérébrales en cas de «troubles». Gall and Spurzheim [1810] ont poussé cette hypothèse plus loin, et ont affirmé que le cortex comprenait des zones spécialisées dans certaines fonctions, alors que la matière blanche constituait le «câblage» reliant ces différentes zones entre elles, une théorie qui a profondément révolutionné les neurosciences. Cette théorie a été vérifiée plus tard par Broca [1861], qui a montré qu'une lésion affectant une région précise du cortex entraînait des troubles du langage. Le paradigme de la spécialisation corticale ainsi que la théorie des neurones récemment introduite a révélé toute l'importance de la matière blanche du cerveau humain dans le fonctionnement du cerveau [Schmahmann and Pandya: 2007a]. Cela a encouragé Dejerine and Dejerine-Klumpke [1895]

et Wernicke à réaliser une description détaillée de la matière blanche dans leurs atlas, et à mettre en relation les changements de structure dans la matière blanche avec des troubles neurologiques. Au cours du XXe siècle, les études concernant la fonction du cerveau ont octroyé une place plus prépondérante encore à la matière blanche. Cette évolution s'est faite en étroite interaction avec le développement de différentes techniques permettant de tracer les faisceaux de la matière blanche [Schmahmann and Pandya: 2007a]. Il a fallu attendre la découverte de l'Imagerie par Résonance Magnétique de Diffusion (IRMD) [Taylor and Bushell: 1985, Websey et al.: 1984a,b, Le Bihan and Breton: 1985, Le Bihan et al.: 1986] et la tractographie pour que l'analyse de la matière blanche puisse être faite *in vivo* et de façon non invasive [Basser et al.: 2000, Mori et al.: 2005, Catani et al.: 2002].

On peut attribuer le succès de l'IRMD à sa capacité à décrire de façon précise la géométrie des structures microscopiques sous-jacentes [Johansen-Berg and Behrens: 2009]. L'IRMD mesure la diffusion moyenne des molécules d'eau, qui est très étroitement reliée à la structure des tissus biologiques, et ce à des échelles bien inférieures à la résolution des images. La diffusion des molécules d'eau est bien décrite par un mouvement brownien quand il n'y a pas d'obstacle dans le milieu ; en revanche, au sein d'une structure fibreuse telle que la matière blanche, les molécules d'eau ont tendance à se déplacer de préférence tangentiellement aux fibres. Grâce à ce phénomène physique, l'IRMD est capable d'apporter des informations sur l'architecture des fibres *in vivo*. L'introduction de la tractographie à base d'IRMD, une technique permettant de retrouver les faisceaux de fibres de la matière blanche, a par ailleurs ouvert de nouvelles perspectives dans l'étude de l'anatomie de la matière blanche, et de ses relations avec certains troubles neurologiques [Ciccarelli et al.: 2008]. Cette thèse se base sur les précédents travaux en IRMD et en tractographie [Basser et al.: 2000, Mori et al.: 2005, Catani et al.: 2002, Behrens et al.: 2003, Lenglet et al.: 2006b, Descoteaux et al.: 2007a]. L'IRMD est non seulement capable de modéliser et d'analyser la structure de la matière blanche à l'échelle du voxel, mais c'est également un puissant outil pour analyser l'anatomie et la connectivité cérébrales. Par ailleurs, cette technique s'est récemment montrée très utile pour l'analyse des anomalies de la matière blanche, et leurs conséquences sur le plan fonctionnel. Le point de départ et la motivation principale de cette thèse a été la publication d'une dissection *in vivo* de la matière blanche, réalisée par Mori et al. [2005] et Catani et al. [2002]. Cette procédure consiste à isoler les faisceaux de fibres de la matière blanche qui jouent un rôle dans une fonction ou une pathologie particulière du cerveau, afin de pouvoir les analyser et les comparer à la recherche de différences caractéristiques. Réaliser cette dissection de façon manuelle est une tâche qui demande une connaissance pointue de l'anatomie cérébrale, et qui peut prendre plusieurs heures. L'élaboration d'une technique permettant d'automatiser l'identification des structures de la matière blanche est donc d'une importance capitale.

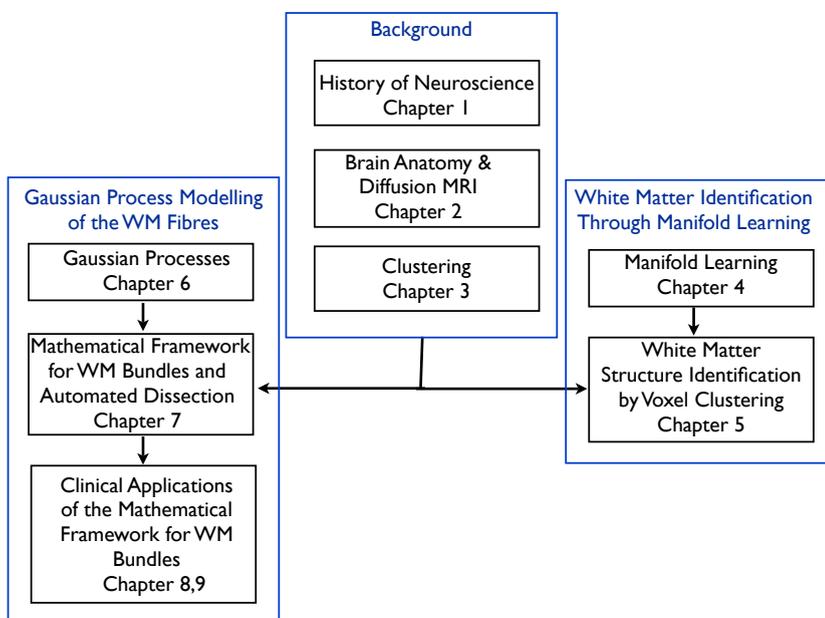


Figure ii.1: Plan des contributions de notre thèse.

PLAN DU MANUSCRIT ET CONTRIBUTIONS

Cette thèse s’articule autour de trois parties. La première partie dresse un historique, puis décrit l’anatomie de la matière blanche du cerveau, le principe de l’IRMD et les techniques et algorithmes nécessaires à la compréhension de cette thèse. La seconde partie décrit ensuite la première contribution méthodologique de cette thèse, à savoir l’identification de structures de la matière blanche par regroupement de voxels en imagerie *QBall*. La troisième partie détaille nos contributions méthodologiques pour le développement d’un outil d’identification automatique des structures de la matière blanche à partir des résultats de la tractographie : nous proposons un cadre mathématique basé sur la notion de processus gaussien pour décrire les faisceaux de fibres de la matière blanche, que nous appliquons à la dissection automatique *in vivo*, au redressement de la moelle épinière et l’étude de différences entre populations de sujets en schizophrénie. L’organisation de cette thèse est synthétisée sur la figure figure [ii.1](#).

Partie I : Rappels généraux

Chapitre 1 donne un aperçu historique rapide des neurosciences, depuis la toute première mise en évidence de la relation entre le cerveau et les tâches motrices et intellectuelles, jusqu’aux développements récents de l’Imagerie par Résonance Magnétique (IRM) qui permettent d’analyser le cerveau *in vivo* et de façon non invasive.

Chapitre 2 répond aux questions suivantes : que cherchons-nous dans la matière blanche du cerveau humain ? En particulier, comment s'organise la matière blanche ? Comment les différentes parties du cerveau sont-elles reliées entre elles ? Quels sont les faisceaux de fibres importants dans le cerveau ? Comment peut-on les analyser à l'aide de l'IRMD ? Ce chapitre regroupe les bases de l'anatomie du cerveau et de la matière blanche, et de son observation grâce à l'IRMD. D'abord, nous rappelons l'organisation générale du cerveau et nous présentons les trois principaux groupes de fibres de la matière blanche, en détaillant parmi chaque groupe les faisceaux les plus importants. Nous introduisons ensuite les principes physiques de l'IRMD, et le lien entre cette imagerie et la structure des tissus biologiques, après quoi nous faisons un bref rappel historique de l'imagerie par résonance magnétique nucléaire de ses débuts jusqu'à la découverte et aux récents développements en IRMD. Le chapitre se termine sur une description de la tractographie basée sur l'IRMD, un outil qui permet de reconstruire les faisceaux de fibres de la matière blanche, ainsi que les applications cliniques principales.

Chapitre 3 répond aux questions suivantes : étant donné un ensemble d'éléments, comment peut-on obtenir de l'information sur chacun des éléments afin de pouvoir les organiser en groupes ? Le but principal de cette thèse est d'identifier les structures de la matière blanche ; pour y parvenir, l'utilisation de techniques dédiées à la séparation des éléments en sous-groupes est fondamentale. Ce chapitre propose un regroupement en familles des algorithmes de classification, et décrit les propriétés représentatives de chaque famille.

Partie II : Identification des structures de la matière blanche par apprentissage de variétés

La seconde partie traite l'identification des structures de la matière blanche, directement à partir des images obtenues en IRMD. A cette fin, nous présentons un algorithme qui a recours à des techniques d'apprentissage de variétés et de classification. La première technique permet de trouver de façon automatique une bonne représentation pour un ensemble d'objets, à savoir les voxels obtenus par IRMD en ce qui nous concerne ; la seconde technique quant à elle concerne la classification de ces voxels.

Chapitre 4 répond à la question suivante : étant donné un ensemble d'éléments, est-il possible de réduire de façon automatique la complexité de leur représentation, en recherchant l'ensemble le plus approprié pour les décrire ? La représentation de la diffusion des molécules d'eau à l'intérieur d'un voxel est un objet complexe. Les algorithmes qui tentent de faire des analyses statistiques des images IRMD doivent compter avec la complexité de cette représentation. Ce chapitre présente un état de l'art des techniques d'apprentissage de variétés. Ces techniques permettent, étant

donné un ensemble d'éléments, d'inférer une représentation de leurs caractéristiques qui est plus synthétique et donc plus simple à manipuler que la représentation originale. Les techniques d'apprentissage de variétés sont au coeur de la contribution présentée dans cette partie, dans le sens où elles permettent de simplifier la représentation des images IRMD, tout en conservant les informations principales.

Chapitre 5 répond à la question suivante : est-il possible de segmenter une image IRMD en plusieurs faisceaux de fibres de la matière blanche de façon automatique ? Ce chapitre présente un nouvel algorithme de classification pour la segmentation des images de diffusion à haute résolution angulaire (HARDI, pour *High Angular Resolution Diffusion Imaging*). Plus exactement, nous comparons dans ce chapitre l'utilisation des images du tenseur de diffusion et des images obtenues en *QBall*, et montrons la valeur ajoutée des images HARDI. Nous montrons également en quoi l'utilisation combinée de techniques d'apprentissage de variétés et de classification permet de proposer un outil bien adapté à l'identification simultanée de plusieurs faisceaux de la matière blanche. Tout d'abord, nous utilisons la relaxation de Markov afin d'intégrer de façon transparente les informations spatiales et les informations de diffusion. Puis la technique d'apprentissage de variétés est appliquée aux cartes de diffusion afin de trouver une représentation adéquate des voxels en IRMD, et dans le but d'améliorer les résultats de la classification. Enfin, un algorithme de classification est lancé sur le résultat de l'étape précédente. Nous présentons des résultats sur des données synthétiques ainsi que sur des images réelles du cerveau humain, afin de mettre en évidence les avantages de cet algorithme.

Partie III : Modélisation des fibres de la matière blanche à l'aide de processus gaussiens

La troisième partie présente la contribution principale de cette thèse : un cadre mathématique pour la description des faisceaux de fibres de la matière blanche. Nous décrivons ici le contexte mathématique, la formalisme que nous proposons et son application à la dissection automatique et *in vivo* de la matière blanche.

Chapitre 6 répond aux questions suivantes : qu'est-ce qu'un outil mathématique approprié pour la représentation des faisceaux de fibres de la matière blanche ? Ce chapitre présente les notions mathématiques nécessaires à la compréhension de la contribution principale de cette thèse. En particulier, en quoi la notion de processus gaussien est un outil mathématique bien adapté à la représentation de fonctions régulières, des courbes et des familles de courbes dans l'espace. Enfin, il présente la façon d'utiliser ce cadre pour analyser de façon quantitative le degré de similarité entre deux fonctions régulières, et comment définir un espace muni d'un produit in-

terne de ces fonctions, de façon à permettre et faciliter l'analyse statistique de ces fonctions.

Chapitre 7 s'intéresse aux questions suivantes : comment peut-on appliquer la notion de processus gaussien à la représentation des fibres de la matière blanche ? Est-il possible d'utiliser l'espace muni du produit interne que définit cette représentation, en combinaison avec des informations anatomiques afin de réaliser une dissection *in vivo* et automatique de la matière blanche ? Dans ce chapitre, nous montrons que le processus gaussien est un modèle bien adapté à la représentation des fibres de la matière blanche. De plus, nous pouvons avantageusement tirer profit du produit interne entre deux faisceaux de fibres représentées comme des processus gaussiens : nous développons un algorithme de classification qui est capable d'incorporer des informations anatomiques. Enfin cet algorithme est appliqué sur une base de données de 68 sujets, ce qui permet d'identifier avec succès les structures les plus importantes de la matière blanche du cerveau.

Chapitre 8 présente d'autres applications du modèle de processus gaussien pour la description des fibres de la matière blanche. Dans ce chapitre, nous proposons une formulation adaptée à la sélection d'une fibre qui soit la plus représentative possible d'un faisceau. Nous utilisons ensuite cette formulation pour développer un algorithme entièrement automatisé de redressement de la moelle épinière, à base de tractographie IRMD ; ceci est à notre sens une application importante. En effet, la visualisation d'images non transformées de la moelle peut rendre le diagnostic compliqué, car la courbure de la moelle ainsi que l'orientation antéro-postérieure peut rendre la visualisation d'une coupe coronale ou une coupe sagittale complète difficile. De plus, cette courbure est très dépendante du sujet, ainsi il est très difficile de faire des études entre sujets au niveau de la moelle épinière. Ainsi, une technique de redressement de la moelle qui permette à chaque image de pouvoir être comparée à une autre constitue une application importante de la tractographie basée sur l'IRMD.

Chapitre 9 montre les utilisations de notre modèle basé sur la notion de processus gaussien afin de mener des études statistiques entre différentes populations. Nous fusionnons les résultats de la classification obtenus grâce à l'algorithme présenté en chapitre 7 sur une base de données composée de sujets sains et de patients atteints de schizophrénie, afin de montrer des différences entre groupes au niveau des faisceaux de fibres de la matière blanche dans le cerveau. En combinant les techniques récentes d'analyse statistiques avec notre cadre mathématique, nous sommes en mesure d'étudier les différences entre groupes de sujets sains et de patients atteints de schizophrénie. Ces études sont menées sur les faisceaux dans leur intégralité d'abord, puis sur une représentation sous forme de squelette des fibres afin de localiser plus précisément certaines différences.

Part I

Background

CHAPTER 1

BRAIN HISTORY

BRAIN, n.: An apparatus with which we think what we think.[...]

The Devil's Dictionary,
Ambrose Bierce, 1911

How long have we been figuring out how the mind and the intellect works? Which were the main paradigms of early neuroscience? When and why did we start thinking that the brain's white matter played an important role in brain function? In this chapter we go through a brief chronicle of neuroscience, neuroanatomy and, in particular the history of the human brain white matter. The chronicle, which is by no means exhaustive, is mainly based on the excellent works of [Gross \[1998b\]](#) and [Schmahmann and Pandya \[2007a\]](#).

1.1 ANCIENT HISTORY: THE SEAT OF THE MIND

Where in the human body is the seat of sensations, feelings and intellect? This question has been around for centuries. The first written reference to the brain, the first hint of its role in intellectual and motor tasks, can be found in the Edwin Smith surgical papyrus (figure 1.1). This papyrus is a surgical treatise originally written approximately 3,000 BCE¹. The publication of an English translation in 1930 [[Breasted: 1930](#)] contradicted previous thoughts on ancient Egyptian medicine which stated that it was a bunch of superstitions and incantations. The manuscript describes diagnosis and treatment in a detailed and methodological manner. Every case is documented in 4 columns: “*title*”, “*examination*”, “*diagnosis*” and, if the physician considered it appropriate, “*treatment*”. The diagnoses are concluded by one of two statements: “*an ailment I will treat*” or “*an ailment not to be treated*”. By analysing the thirty four cases, the fact that the surgeon seemed to be aware of the cause-consequence relation between head injury and behavioural change clearly stands out. For instance, it is possible that the following transcription of the trans-

1. In this chapter we use historical writing conventions for the dates: BCE means Before Christian Era, CE means Christian Era and ca. stands for *circa* or “approximately in”.

lated papyrus reveals the surgeon relating a brain injury with aphasia [Wilkins: 1964]:

“ **Case Twenty**

Title Instructions concerning a wound in his temple, penetrating to the bone, (and) perforating his temporal bone.

Examination If thou examinest a man having a wound in his temple, penetrating to the bone, (and) perforating his temporal bone, while his two eyes are blood shot, he discharges blood from both his nostrils, and a little drops; if thou putttest thy fingers on the mouth of that wound (and) he shudder exceedingly; if thou ask of him concerning his malady and he speak not to thee; while copious tears fall from both his eyes, so that he thrusts his hand often to his face that he may wipe both his eyes with the back of his hand as a child does, and knows not that he does so...

Diagnosis Thou shouldst say concerning him: “One having a wound in his temple, penetrating to the bone, (and) perforating his temporal bone; while he discharges blood from both his nostrils, he suffers with stiffness in his neck, (and) he is speechless. An ailment not to be treated.”

Treatment Now when thou findest that man speechless, his [relief] shall be sitting; soften his head with grease, (and) pour [milk] into both his ears.

”

Despite the existence of the Edwin Smith papyrus, the Egyptians considered the brain a minor organ, placing the seat of the mind and the soul in the heart. The funerary rituals described in the *book of the dead* can be taken as evidence of this. During the rituals the embalmers extracted the heart with great care and weighed it against feathers to determine its tendency towards good and evil. The brain, on the other hand, was extracted through the nostrils and discarded [Faulkner: 1972]. Although the belief that the heart was the seat of perception and cognition seems to have been common along the pre-Columbian, Babylonian, Indian and Mesopotamian cultures [Obeyesekere: 1977, Roux: 1966], there are some exceptions like the Incas who practised trepanation as a cure for some sorts of mind-related disorders [Jørgensen: 1988, Sosa and Valderrama-lturbe: 2001].

Three Greek medical schools

Centuries later, in Greece, new arguments for the brain being the centre of sensation and intellect emerged. The first Greek writings stating this hypothesis belong to Alcmaeon of Croton (ca. 450 BCE). Alcmaeon, a prominent member of the medical

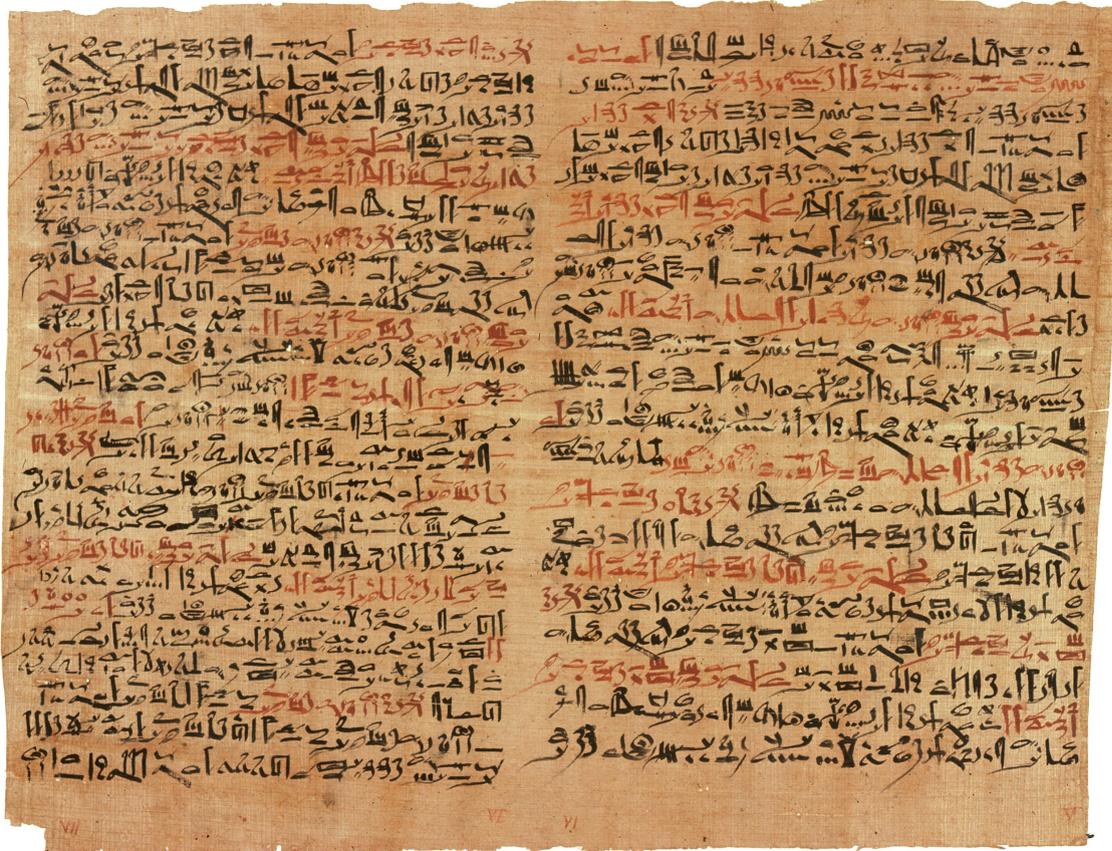


Figure 1.1: Edwin Smith Papyrus in its original version, written in a hieratic script

school of Croton, seems to have been one of the first practitioners of dissection as a way to look for anatomical knowledge. Through dissection studies, he developed detailed theories on the senses, with a particular interest in vision. Alcmaeon described the optic nerves through their crossing “behind the forehead”, becoming the author of one of the first written documents identifying a major white matter structure. Moreover, he used this evidence to state a hypothesis on the coordinated movement of the eyes.

Even though the doctrine of Alcmaeon was shared by several medical schools and Greek physicians such as Hippon of Samos (ca. 470 BCE) and Anaxagoras of Calzomenae (ca. 428 BCE), this view was not dominant in the Greek world. Empedocles (ca. 430 BCE), taught that the blood was the vessel of all sensations and thoughts and saved a prominent place for the heart concerning those functions. Being a leading member of the medical centre at the second most important centre of Greek medicine, Agrigentum, he was an influential figure of his time.

Not far from there, on the island of Cos, Hippocrates (ca. 425 BCE) exercised his role as a physician and philosopher at the first leading centre of Greek medicine. Unlike the school of Cos, Hippocrates and his pupils were not fond of dissection as a means to expand their medical knowledge. Their comprehension of the body and

its functions was based on external observation and behavioural changes specific to different diseases. The Hippocratic texts are written in different styles and with various levels of detail which leads us to believe they were written by a Hippocratic society rather than by a single individual. Within these texts, the brain is considered the origin and cause of all intellectual and sensory functions:

“ It ought to be generally known that the source of our pleasure, merriment, laughter, and amusement, as of our grief, pain anxiety and years is none other than the brain. It is specially the organ that enables us to think, see, and hear, and to distinguish the ugly and the beautiful, the bad and the good, pleasant and unpleasant [...] ”

Against observational science

Even though he was not fond of the pre-Socratic tradition of natural philosophy, Plato (427-347 BCE) also championed the brain as the seat of the intellect. Through the foundation of his philosophical school in Athens, The Academy, and his written dialogues he shaped philosophy, law, politics and science as we know them today.

Plato disregarded empirical exploration of the universe, considering the observable part of everything a mere superficial appearance, an imperfect and transient illusion of the sensory world. According to his teachings, a philosopher should understand the cause and consequences of each kind of object through abstraction and introspection. This concept was strongly established in the seventh volume of Plato's masterpiece, *The Republic*, known as the *allegory of the cave*: “The philosopher must escape the tyranny of sensory experience and empirical knowledge and climb out of the cave in order to reach the higher realities of true knowledge” as synthesized by Gross [1998b]. His views of the brain were exposed in detail in the *Timaeus*, an essay of great influence in the Middle Ages. According to it:

“ It is the divinest part of us and lords over all the rest. ”

Plato gave a strong structure to philosophical, political and governmental systems which still holds. Despite that, through the *Timaeus*, his disregard of empirical science was deeply influential during Medieval times. His teachings therefore have been referenced as “essentially evil” and “an aberration” by modern science historians [Lloyd: 1970, Sarton: 1993].

Plato's prodigal student

Although Plato advocated “against” empirical science, a student who spent twenty years at the Academy is inscribed into history as one of the greatest naturalists of all times. Aristotle was born in 384 BCE to a medical family. His family, well positioned in the Macedonian kingdom, and his reputation led to the appointment of Aristotle as tutor of Alexander the Great. Funded by Alexander, Aristotle was able

to found his school and research centre, the Lyceum, and to endeavour numerous philosophical and natural research programs.

Aristotle is considered the first comparative biologist and one of the greatest naturalists of all times. Despite his awareness of the previous work of Alcmaeon, Hippocrates, Plato and others, he was one of the most fierce supporters of the heart as the seat of the mind. Several of his writings strongly advocated against the brain as the centre of sensation

“ And of course, the brain is not responsible for any of the sensations at all. The correct view [is] that the seat and source of sensation is the region of the heart. ”

He relegated the main functions of the brain to the cooling of the blood. A hypothesis which he upheld through several arguments, some of them based on the consistence of the brain:

“ For if the brain be either too fluid or too solid, it will not perform its office, but in the one case will freeze the blood and in the other will not cool it at all, and thus, cause disease, madness and death. ”

Aristotle meticulously dissected 49 animals from elephant to snail and no human beings. He was able to notice that their circulatory systems were more spread along the body than the nerves. Most of his arguments are summarized in table 1.1, adapted from Gross [1995]. This doctrine was not supported by Aristotle alone, Diocles of Carystus (fourth century BCE) and Praxagoras of Croton (ca. 340 BCE), also asserted the heart as the centre of intellect and sensation [Crivellato and Ribatti: 2007]. Both physicians considered the heart, arteries and veins as a pumping system for the *physic pneuma*, the light and invisible substance believed to perform sensory, motor and intellectual activities.

Despite his heart advocacy, Aristotle greatly influenced the subsequent development of neuroscience. Being the most influential naturalist until the late Middle Ages, his stress on dissection encouraged several scholars to use this technique as a principal resource for research. More directly, he contributed significantly to the creation of the Museum at Alexandria, where systematic human neuroanatomy emerged. Regarding his (mis)interpretation of brain function, perhaps the key was in the absence of clinical observation, central in the Edwin Smith surgical papyrus and on the studies of Alcmaeon, the Hippocratic doctors and other ancient scholars.

The Museum at Alexandria

The Museum at Alexandria was founded by Ptolemy I, a general of Alexander the Great who became the first Greek ruler of Egypt by the end of the fourth century BCE. Being a student of Aristotle, he was in close ties with the Lyceum and fond of philosophical activities. Ptolemy commissioned the conception of the museum to

Brain	Heart
Affected by emotion	Not affected
All animals have a heart or similar organ	Only vertebrates and cephalopods have one, and yet other animals have sensations
Source of blood which is necessary for sensation	Bloodless and therefore without sensation
Warm, characteristic of higher life	Cold
Connected with all the sense organs and muscles via the blood vessels	Not connected with the sense organs or the connection is irrelevant
Essential for life	Not so
Formed first and last to stop working	Formed second
Sensitive	Insensitive: if the brain of animal be laid bare, it may be cut without any signs of pain or struggling
In a central location, appropriate for a central role	Not so

Table 1.1: Comparative table of Aristotle's analyses of brain and heart. Adapted from [Gross: 1995]. The table is a compilation of citations from several works of Aristotle [1942, 1955, 1957a,c,b].

Demetrius of Phaleron (ca. 350-280 BCE) and Strato of Lampsacus (ca. 335-269 BCE). Both of these Greek philosophers had studied at the Lyceum. Strato even became the third head of the philosophical school. The three of them conceived a state-supported research institute where over a hundred professors were living on site with their salaries and expenses paid. The Museum had lecture and study rooms, a botanical garden, a zoo, an astronomical observatory and dissecting and operating rooms.

Herophilus (ca. 270) and Erasistratus (ca. 260), founders of the systematic study of the structure of the human body [Von Staden: 1989], worked within the walls of these operating rooms. Being in Alexandria, the dissection of human bodies was simpler than in ancient Greece as Egyptians did not worship the integrity of the dead human body in the same way Greeks did. Moreover, the custom of mummification probably simplified the acceptance of "working" with a dead human body. A minor extra detail that might have helped the two physicians to work freely with human bodies was the complete support of the foreign rulers of Egypt, the Macedonian dynasty of Ptolemies. Among the licences from the rulers of Egypt that Herophilus and Erasistratus enjoyed was the authorization to perform live dissection of human beings. The practice of human vivisection is not documented in any other place or epoch except for the German and Japanese research programs of the second World War. Furthermore, the dissection of human cadavers as a means to study the structure of the body in the Western civilization disappeared the thirteenth century [Gross: 1998b].

The study of the brain was of particular interest to Herophilus and Erasistra-

tus. They believed it has a dominant role in sensations, thought and movement. The first detailed description of the brain including the ventricles was an outcome of their research². Herophilus located intelligence in the brain, was the first to produce anatomical drawings and to distinguish sensory and motor nerves. Erasistratus, located the origin of the nerves in the brain, stated that the fourth ventricle was the “command centre” of the brain and likened the gyri of the brain to the coils of the small intestine. This intestinal metaphor was sustained until the nineteenth century, a fact that can be appreciated in many anatomical drawings where the appearance of the brain is similar to that of the small intestine. Erasistratus compared the human brain with the brains of several animals and claimed that the intelligence of the human being was due to the larger number of gyri. Galen ridiculed this statement by drawing attention to the convoluted brain of the ass. This “joke” was often cited during the following 1,500 years after, and seems to have inhibited any serious study on the cerebral gyri until the seventeenth century. Also, Erasistratus performed experiments on the living brain, unfortunately no document of these studies survives. After the period of the two physicians, medical studies in Alexandria decayed rapidly, turning to a more arcane practice.

Galen of Pergamon

Galen (129-199 CE), was the most influential figure of ancient medical science. History, physiology, anatomy and medicine were shaped after Galen’s manuscripts from medieval times until our days. Central to Galen’s doctrine, was his belief in the absolute need for a rigorous anatomical methodology. He claimed, in fact, that only correct dissections would provide “apodeictic proofs”, or incontrovertible demonstrations, enabling the researcher to draw legitimate conclusions.

Galen took an interest in the brain functions considering that

“ The brain is that part which receives all sensations, produces images and understands thoughts ”

He fought the contemporary stoic doctrine that placed the heart at the centre of the intellectual and sensory system by providing arguments based on rigorous anatomical demonstrations. In one of these demonstrations, he proved that the voice disappears after an incision in the inferior laryngeal nerves is made in the pig [Gross: 1998a] and other animals. His research yielded a detailed description of the brain. However, the accuracy of his work was not appreciated until recent historical analyses stated that the brain he described probably belonged to an ox, which in Galen’s time was more easily available than the human brain.

Galen provided a detailed map of the brain. He paid particular attention to the ventricles and described them in detail. Most probably, he was following the program of Erasistratus, where the ventricles had a crucial physiological role. He

2. For the readers not familiarized with brain anatomy, an introduction to it is given in chapter 2

linked these programs and classical Greek conceptions of the soul by stating that the ventricles, being a cavity, were the harbour of the substance performing all the intellectual and sensory functions, the *physic pneuma*. He had a mechanic view of brain function. The *physic pneuma* was produced by mixing *vital pneuma*, the inspired air that had already passed from the lungs and came from fine capillaries, with external air coming from the nostrils. This mixing was produced within the anterior and middle ventricles and at the base of the brain. Galen's point of view persisted during thirteen centuries [Crivellato and Ribatti: 2007]. Despite this important role assigned to the ventricles, he located the seat of the soul in the brain tissue, a conclusion supported by the theory that the brain was the only source of the spinal cord and other nerve systems in the body.

Around Galen's death, in 199 CE, Greek science and medicine went into decline. Analysis and study were replaced by dogma and prayer, and exorcism took the place of medical healing.

1.2 MIDDLE AGES: SMALL ROOMS ---

During the Middle Ages, the main trend in the Western world with respect to brain research was about the localization of mental faculties. Following Galen's doctrine, these faculties were localized in the ventricles. The brain was divided into three small rooms or cells. The first one comprising the two lateral ventricles, received the input from all the sense organs and was the site of the common sense or *sensus communis*, that integrated across modalities. The second, or *middle*, cell, was responsible for the cognitive processes: reasoning, judgement and thought. The third cell, or *ventricle*, was the site of memory. The Persian philosopher Ibn Sina (ca. 980-1037), also known as Avicenna, documented this doctrine in one of his many medical writings. This can be observed in a 1347 edition of his work *De generatione embryonis* shown in figure 1.2.

Even though Galen postulated that the intellectual faculties were located in the solid sections of the brain, with the advent of Christian philosophers a relocation to the ventricles took place. Nemesius, bishop of Emesia (ca. 390) and St Augustine (350-430), supported the non-material nature of the soul with the cell doctrine [Tascioglu and Tascioglu: 2005]. According to this doctrine, the tissue was too earthy to contain the soul and that its seat were the "empty" cells of the brain. The three cells were thought of as the three sections of a temple, or the three parts of the holy trinity. Hieronymus Brunschwig (1450-1512) illustrated this in coloured woodcuts in his 1512 work *Liber de Arte Distillandi*, a practical manual on chemistry, alchemy and the distillation devices used to manufacture drug therapies, see figure 1.3.

It was not until the tenth century that the main doctrine changed. The cell principle remained intact during 1,200 years with only minor developments [Pagel: 1959]. Then, the process was conceptualized as a more dynamic one by presenting an anal-

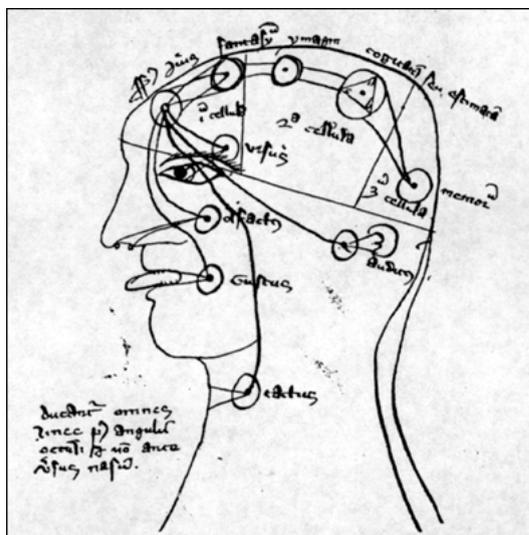


Figure 1.2: Ventricular theory of sensation and perception illustrated in a Latin edition of *De generatione embryonis* edited in 1347 and written by the Persian philosopher Ibn Sina between 980 and 1037.

ogy with the digestive system: the first cell received the sensory inputs and made them into images; these images were warmed up in the middle cell making them more appropriate for further processing into cognition; finally, the mental leftovers were stored in the third cell.

1.3 RENAISSANCE: THE HUMAN BEING IS THE CENTRE _____

The Renaissance period brought the human being back to the focus of the Western world. Art and science were again centred on the details of the human body. In the arts, the muscles and factions were painted and sculpted with exquisite detail; and in medical sciences and natural philosophy of the human being, anatomy and dissections were, once again, the purpose and main resource for intellectual inquiry. Along with more detailed anatomical studies, illustrations were becoming more sophisticated in the hands of Albrecht Dürer (1471-1528) and Leonardo da Vinci (1452-1519).

Leonardo Da Vinci, performed several studies on human anatomy and function, ranging from intercourse, to the optics of the eye. Regarding the anatomy and functions of the brain, he studied the ventricular and optical systems as shown in figure 1.4. By injecting hot wax into the brain, he was able to extract a model of the ventricles and study its shape. Moreover, he performed a meticulous study of the nerves connecting the eyes, nose and mouth to the brain, paying particular attention to the optical tract as it can be observed on the right plate of figure 1.4.

In Padua, Andreas Vesalius (1514-1564) reinvented anatomy, by daring to challenge the dominant doctrine originated in the work of Galen. Vesalius is often paired with Copernicus as an initiator of the scientific revolution. His masterpiece, *De Hu-*

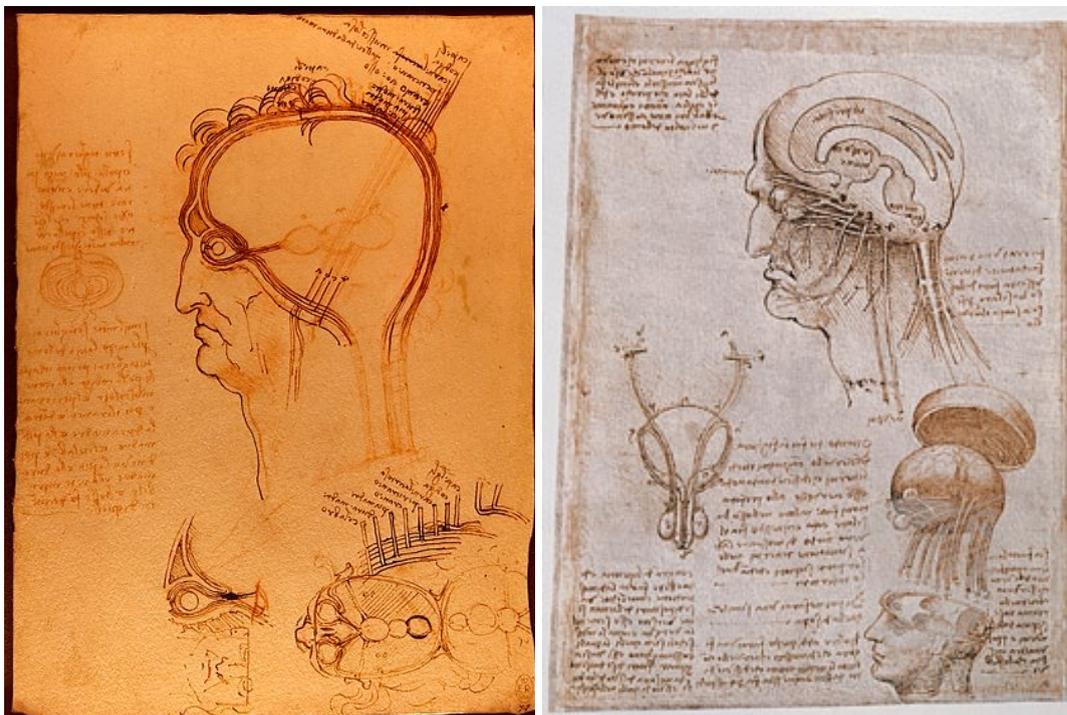


Figure 1.4: Representation of the ventricular system and of the brain anatomy by Leonardo Da Vinci.

Like Galen, he rejected Erasistratus' doctrine on gyri. However, he advocated for the importance of the gyri as a way for the blood vessels to nourish the inner parts of the brain and stated that they should be studied:

“ [You] may learn the shape of this twistings by observing the brain of some animals [on your plate] at breakfast or at dinner ”

Furthermore, he was the first to distinguish the grey matter calling it *cerebrum* and white matter, the *medulla*. Vesalius used different anatomical markings for the cortex and the white matter in his drawings (see for instance the right plate of figure 1.7). Remarkably, he observed that the *medulla* was continuous with the corpus callosum, named in this way due to its particular hardness, and that this structure links the two halves of the brain. This division was confirmed and studied in depth by Archangelo Piccolomini (1526-1586), professor of anatomy in Rome, who succeeded in separating the grey and white matter through gross dissection and noting “certain lines” in the cerebrum.

Up to this point the analysis of the grey and white matter of the brain underwent the same historical path. However, from this historical point and on we could focus on the functional or anatomical study of the brain. In the remainder of this chapter we will focus only on the history of the white matter anatomy and function, as its analysis is the main subject of this thesis.



Figure 1.5: Cover artwork from the 1543 edition of *De Humani Corporis Fabrica* by Andreas Vesalius, who is shown performing an anatomical dissection. It is believed that it was drawn by one of the painters working at the school of Titian.

1.4 MODERN AGE: GROSS TRACTOGRAPHY

The method of gross dissection of the human brain's white matter was developed by Arcangelo Piccolhomini (1526-1605) in 1586. This dissection technique enabled scholars to study the white matter in depth and elaborate theories of brain function which considered the cortex and the white matter as having separate but complementary roles in brain function. Among them, Thomas Willis (1621-1675), one of the most important medical figures after Galen, stated that the cortex was involved in memory and will. The sensory stimuli came through medullary pathways to the corpus striatum, became perceptions and imagination in the corpus callosum and then stored as memories in the cortex. He upheld that the cortex was the initiator of voluntary movement and the cerebellum of involuntary movement. However, Willis conceived the cortex as a "single organ" with a single function. This is observable in his ventral drawing of the brain in figure 1.8, where there is a fine-grained labelling of the cranial nerves and basal structures but the whole cortex is labelled with a single "A". Being a medic by profession, most of his conclusions were based

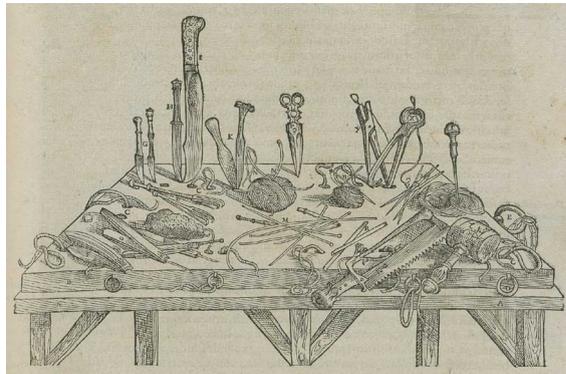


Figure 1.6: Artwork from the 1543 edition of *De Humani Corporis Fabrica* (1543) by Andreas Vesalius, showing the instruments used for dissection.

on observations of human pathologies and experiments with live animals. These studies led him to notice the great variability of the gyrus patterns among different species and to correlate this variation with intellectual capabilities:

“ Hence, these folds or gyri are far more numerous and rarer in man than in any other animal because of the variety and number of acts of the higher Faculties, but they are varied by a disordered and almost haphazard arrangement so that the operations of the animal function might be freely changeable and not limited to one. ”

Willis thought of the brain and the white matter in terms of the Galenic paradigm, and assigned to the white matter the role of a highway for animal spirits to flow from the cerebrum and cerebellum into the nervous system. Spirits directed to the brain carried the sensations and the ones flowing outwards from it served as carriers of locomotor function.

The founder of microscopic anatomy, Marcello Malpighi (1628-1694), professor in Bologna, was the first to examine the cortex at such a small scale. Within his studies, he observed a set of glands with tubes perpendicular to the cortex, as shown in figure 1.9, one of the illustrations from his work *De Cerebri Cortice* (1666). Being a dedicated botanist, Malpighi found parallels between plant and the structure of the brain:

“ [These fibres] ramify from four reflected crura of this medulla in all directions, until they end by their branched extremities in the cortex ”

Even though his drawings and observations seem to fit the description of pyramidal neurons, it is likely that these were artefacts of his dissection and analysis process. This is believed to be the case because of his written observations stating that the “glands” were more prominent in boiled than fresh tissue [Gross: 2007]. The glandular theory of the brain that Malpighi upheld was common in the seventeenth and eighteenth centuries and conformed to one of the two most important Greek doctrines

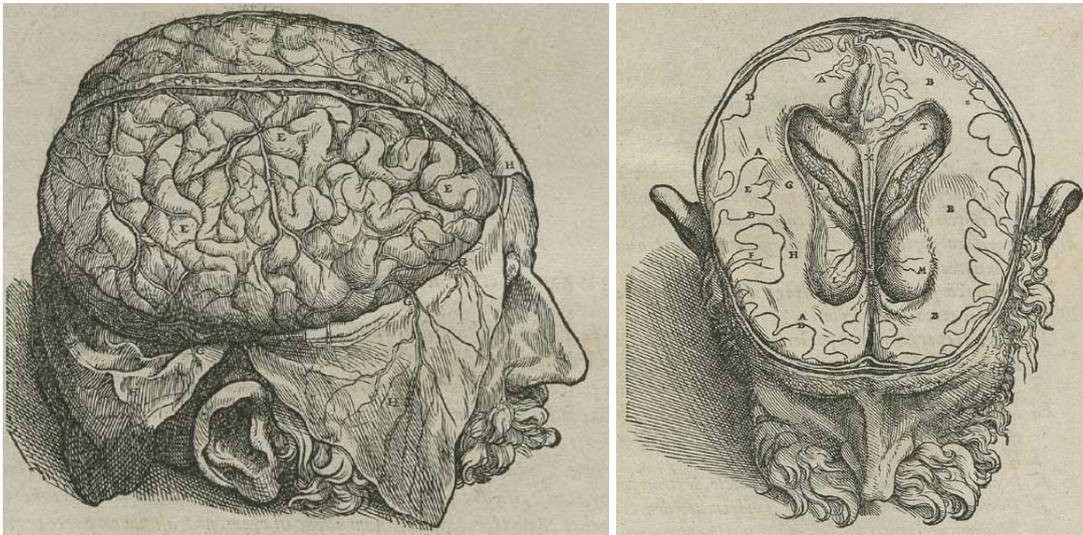


Figure 1.7: Even if his texts stated that more attention should be paid to the gyri of the brain, the drawings of Andreas Vesalius show that his dissection studies were more focused on the ventricles, blood vessels and meninges and distinguished the grey matter from the white matter. This can be observed in the difference of detail in his drawings from *De Humani Corporis Fabrica* (1543).

of the brain, the Aristotelian. This doctrine claimed that it was a cooling organ and the Hippocratic postulating it as the source of the bodily humour responsible for introspection and solid temperament, the phlegm. At the same time, the opposed doctrine, supported by Fredrik Ruysch (1628-1731) and Thomas Bartholine (1660-1680), professors in Amsterdam and Copenhagen respectively. This doctrine advocated that the cortex was vascular, made almost purely of blood vessels and that the gyri were mechanisms for protecting such a delicate vascular structure.

Based on the observations of Malpighi, Nicholaus Steno (1638-1686), argued that in order to discover the purpose of these fibres, their pathways should be traced through scraping and studied [Clarke: 1970]. In doing so, he started a long debate on whether the right way to trace fibres was from the cortex to the stem, or the other way around. These heated debates seemed to be customary at that time in Europe, as it was illustrated using eggs by Swift [1726]. Willis implemented Steno's idea and showed that the white matter was composed of an intricate arrangement of fibres grouped in bundles. Using the scraping technique as a main tool of inquiry, Raymond de Vieussens (1635-1715) and several others studied the structure and distribution of the white matter bundles [Schmahmann and Pandya: 2007a]. Particularly, Felix Vicq D'Azyr (1748-1794) differentiated association from commissural fibres and studied bundles in detail several. He proposed that the multiple connections had a role in preserving the normal functioning of the brain in the case of "inconveniences".

By the beginning of the nineteenth century, there was a consensus on the complexity and heterogeneity of the white matter. Several methods to perform its fixation and analysis through gross dissection were available. The white matter bun-

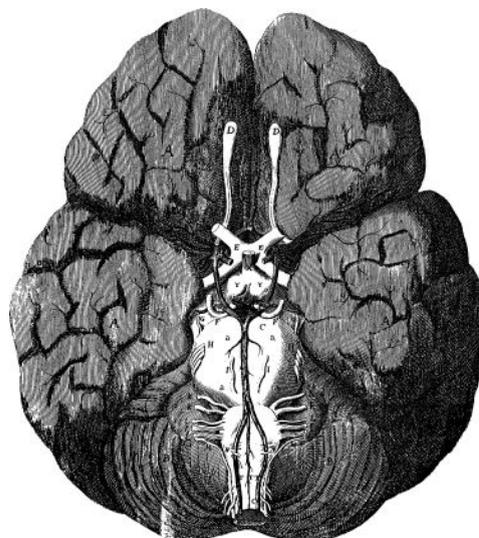


Figure 1.8: Illustration from *Cerebri Anatomie* (1664) by Thomas Willis. Willis conception of the cortex as a “single organ” with a single function, in this illustration it is noticeable the fine-grained labelling of the cranial nerves and basal structures contrasting with the whole cortex, labelled with a single “A”.

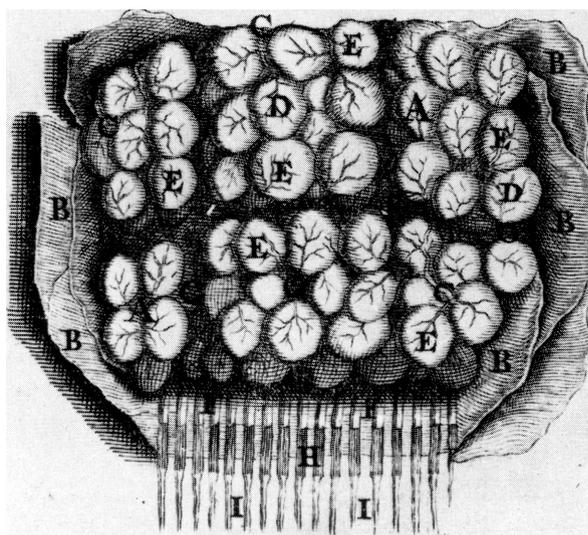


Figure 1.9: Cortical “glands” as illustrated by Marcello Malpighi in his *De Cerebri Cortice* (1666). The illustration seems to be depicting pyramidal neurons and their axons, perpendicular to the cortex and going into the white matter. However it is likely that he was observing artefacts. This is believed to be the case because Malpighi reported that the glands were more prominent in boiled than fresh tissue.

dles were divided in association and commissural groups and many of them were being studied in detail. Researchers like Herbert Mayo (1796-1852), Friedrich Arnold (1803-1890), Karl Burdach (1776-1847), Louis-Pierre Gratiolet (1815-1865), Achille Foville (1799-1878) and Johan Reil (1759-1813) among others, produced a large corpus of studies on the bundles of white matter based on gross dissection and microscopy. Several of these having exquisite illustrations and remarkable discoveries



Figure 1.10: Illustration from *Traite d'Anatomie et Physiologie* (1786) by Félix Vicq d'Azyr. The gyri are drawn resembling the small intestine, following the doctrine of Erasistratus.

about new bundles and the organization of the human brain's white matter [[Schmahmann and Pandya: 2007a](#)].

Concerning the cortex, Thomas Willis, François Pourfour du Petit (1644-1741) and Emmanuel Swedenborg (1688-1772) had elaborated theories assigning it a fundamental importance in sensory and motor functions [[Gross: 1998b](#)]. Still, the predominant doctrine considered it to be either vascular or glandular, with a principal function of nourishing the subcortical structures or cooling down the blood. Even more, most of the illustrations until the mid-eighteenth century showed the cortex and its gyri like the intestine, following Erasistratus, and no effort was taken in performing an accurate depiction of the gyri as shown in figure 1.10.

Cortical specialization

[Gall and Spurzheim \[1810\]](#) shed new light onto the conception of the cortex. They stated it had areas with specialized functions and the white matter was the “wiring” connecting them, a theory that revolutionised neuroscience. The illustrations performed and included by Gall and Spurzheim showing the anatomy of the human brain, along with several animals, was superior to any of the studies previously published [[Monro: 1813](#)]. In their work, they studied the white matter in detail and added the category of projection fibres, which connect the brainstem and the cortex, to the categorization proposed by Vicq D’Azyr. They claimed that by studying the shape of a person’s head, several intellectual and moral dispositions could be identified. They called this field of research “Phrenology”. This claim, although extensively supported in their four-volume publication, led to a heated debate around their work and provoked the most critical responses to their work.

Aside from the claim about the inference of intellectual disposition by analysing the shape of the head, the cortical specialization paradigm has been repeatedly proven since Gall and Spurzheim first published their work. One of the most significant demonstrations was provided by [Broca \[1861\]](#). Broca presented the brain of a recently deceased patient who could only say “tan” to the medical society of Paris. The brain of this patient was damaged in a specific area of the frontal lobe, which now bears the name of Broca. Over the next few months, he displayed the brains of several patients with similar pathologies and a lesion in the same area. Through this, Broca demonstrated that damage in a specific area of the cortex caused speech impediments, proving cortical specialization.

1.5 NINETEENTH CENTURY: CHEMICAL TRACTOGRAPHY _____

The gross dissection method proved to be extremely useful in the first era of white matter analysis, however due to its technical limitations, it was not capable of resolving certain questions. Firstly, at that scale of analysis it was not possible to determine the direction of the fibres; the distinction of afferent from efferent connections was impossible to achieve. For this reason, it was unfeasible to tell if, for instance, projection fibres were going to the cortex, emanating from the cortex or just turning around and coming back to the brainstem. By the middle of the nineteenth century, the study of white matter had come to a point where microscopy was needed to further elucidate its characteristics [[Todd: 1845](#)].

Parallel to the gross dissection-based analysis the neuron theory was developed during the late seventeenth and early nineteenth centuries. By the end of the eighteenth century, Evangelista Purkinje (1787-1869) and Theodore Schwann (1810-1882) among others, had described the nerve cell and the axon [[Schmahmann and Pandya: 2007a](#)]. However, it was not until the development of optical physics led to the achromatic microscope, and chemistry led to staining techniques, that neurons could be visualized and studied in detail. This was particularly achieved by the staining method developed by Camillo Golgi (1843-1926) first published in 1883 [[Golgi et al.: 2001](#)]. Santiago Ramón y Cajal (1852-1934) used Golgi’s technique extensively. Through this precise staining, he was able to clearly illustrate the shape of the neuron and emphasize the structural, functional and developmental singularity of the nerve cell, which is now known as the “neuron doctrine”. Golgi supported a diametrically opposed view: called the “reticular doctrine” in which the neurons were not singular elements but just nodes on a continuous complex network that performed all the processing of the nervous system. Cajal and Golgi jointly received the Nobel Prize in medicine for their work in 1906. Presently, the “neuron doctrine” has prevailed and the works of Golgi and Cajal, remain fundamental for neuroscience [[Glickstein: 2006](#), [Bullock et al.: 2005](#)].

The advances in microscopy and staining in the first half of the nineteenth century which led to the acceptance of the “neuron doctrine”, also enabled white matter

by clinicopathological cases. Based on the paradigm of cortical specialization and on the knowledge of affected cortical areas on a number of pathologies, he studied which pathways were connecting these areas to the rest of the brain and drew conclusions about different cortical systems implicated in these disorders. Through the study of the relationship between pathology and structure and his historical accounts, Dejerine attempted to resolve the existing discrepancies regarding the structure of the white matter. His numerous and fundamental contributions to clinical neuroscience and to the study of brain anatomy are difficult to summarize.

Clinical neurology was emerging as a discipline at the time that Dejerine and Wernicke published their studies. At that time, Jean-Martin Charcot (1825-1893) had been leading a series of studies linking white matter lesions to its clinical consequences for several years. The complexity of the white matter was perceived at a completely different scale than ever before; the intricacy of its pathways and the importance of the different types of cortical area linkages were recognized as subjects of great importance [[Ramón y Cajal: 1933](#)]. However, the study of these cortical connections, recognized as a fundamental issue in understanding the brain in depth, had not been performed yet due to technical limitations.

Chemical tracing by staining combined with histology, still persists as the principle technique to investigating cortical connections. The development of anterograde tract tracers capable of trace the connections from the neuron to the axon and retrograde tracers capable of tracing from the axons to the cell bodies facilitated the detailed study of cortical connectivity [[Schmahmann and Pandya: 2007a](#)]. These techniques enabled researchers to differentiate axons leaving from a cortical area, from the ones arriving to it. Today, there is an availability of chemical tracers which are capable to stain axons individually, enabling a precise study of cortical connections in the brain. However, a major distortion in these type of studies is due to the histological process, in which the brain must be excised, fixated and sectioned in order to contemplate the results of these techniques.

1.6 TWENTIETH CENTURY: IMAGING-BASED TRACTOGRAPHY .

With the purpose studying the cortical connections without the need for histology, [Cowan et al. \[1972\]](#) developed autoradiographic tracing. This technique consists in the injection of a radioactive chemical tracer is used in combination with radiography to visualize axonal pathways. In addition to the visualization of nerve endings, this technique also displayed the course of the axons from the injection site to its distant terminations. This technique consists in the injection of a radioactive chemical tracer is used in combination with radiography to visualize axonal pathways. In addition to the visualization of nerve endings, this technique also displayed the course of the axons from the injection site to its distant terminations. The introduction of this technique set the grounds for a considerable amount

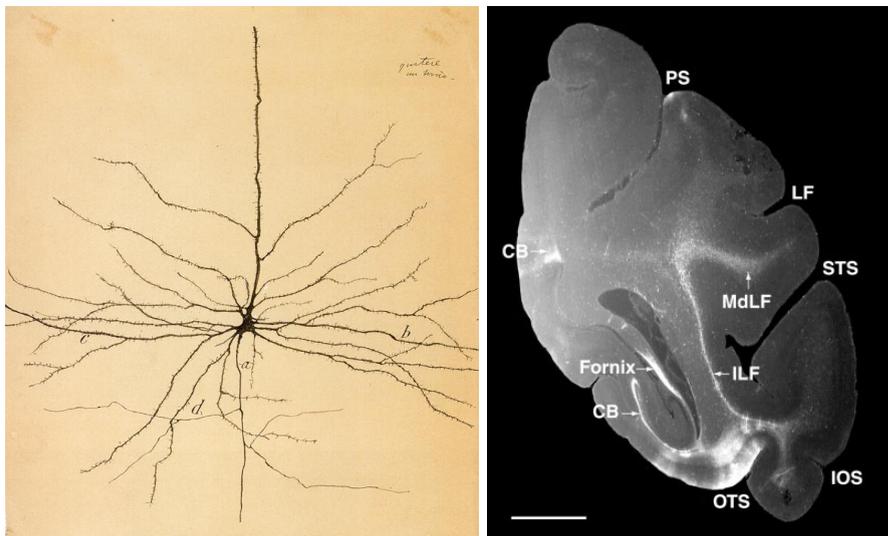


Figure 1.12: Chemical tracing, from Golgi staining to autoradiography. On the left a pyramidal neuron reproduced from the original histological atlas of Ramón y Cajal [1995]. On the right autoradiographic tracing reproduced from Schmahmann et al. [2007]. In this radiography the injection site can be noted close to the OTS label and several white matter structures have been labelled.

of research which investigated several characteristics of the brain such as: cortico-cortical connections, subcortical connections and the delineation of hemispheric fibre systems [Schmahmann and Pandya: 2009]. A comparison between the first axonal tracing techniques, performed by Ramón y Cajal, and state-the-art autoradiographic tracing performed by Schmahmann et al. [2007] is shown in figure 1.12. Autoradiographic tracing nonetheless, is suited to be used on experimental animals only. The injection of radioactive substances directly on the brain makes this invasive technique impossible to be applied for the study of the human brain's white matter *in vivo*.

In introducing diffusion weighted magnetic resonance imaging and then diffusion tensor imaging, Websey et al. [1984a,b], Le Bihan and Breton [1985], Taylor and Bushell [1985], Le Bihan et al. [1986] and Basser et al. [1994b], set the grounds for *in vivo* tractography of the human brain. These techniques rendered possible to measure the directionality of diffusion in the whole brain, divided in millimetric sections, *in vivo* and non-invasively. Using these measures, it was feasible to trace the white matter fibres either in the shape of three-dimensional trajectories representing small bundles [Basser et al.: 2000, Mori et al.: 2005, Catani et al.: 2002] or in terms of the probabilistic measure of connectivity between areas of the brain [Behrens et al.: 2003, Anwander et al.: 2007, Descoteaux et al.: 2009a]. Both types of tractography studies are shown in figure 1.13.

1.7 AUTOMATED *In Vivo* DISSECTION OF THE WHITE MATTER

Today, the white matter fibre bundles can be estimated and traced *in vivo* through

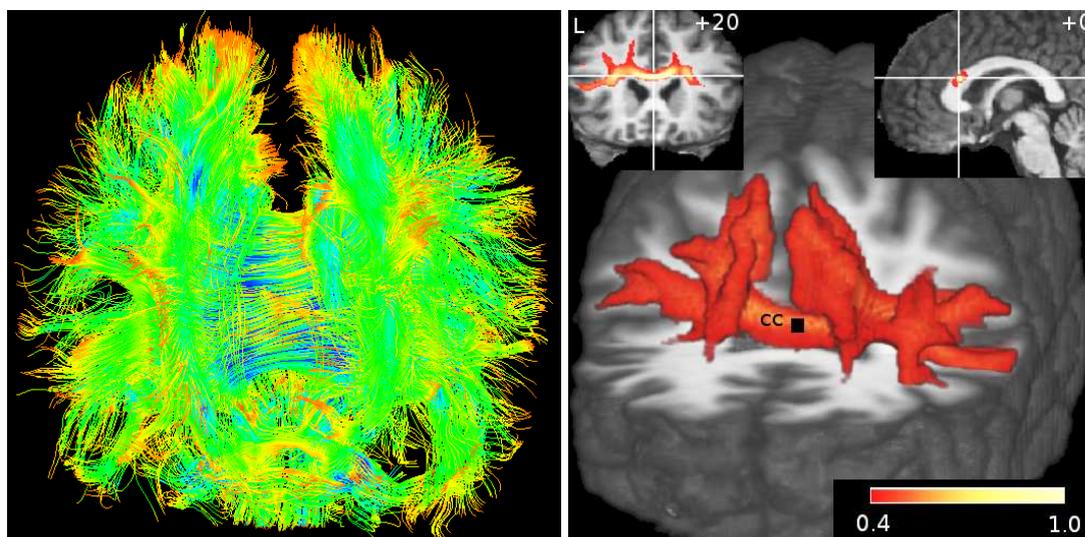


Figure 1.13: Diffusion-based tractography studies. On the left, streamline tractography results obtained using the techniques proposed by [Basser et al. \[2000\]](#). On the right, probabilistic tractography results reproduced from [Descoteaux et al.: 2009a](#).

Diffusion MRI-based tractography techniques. The works done by [Catani et al. \[2002\]](#) and [Mori et al. \[2005\]](#) show that by using a meticulous procedure, somewhat similar to the procedure of gross dissection, it is possible to effectively recover major white matter structures with impressive precision. Then, works like the ones done by [Lawes et al. \[2008\]](#) and [Dauguet et al. \[2007\]](#) have validated against gross dissection and autoradiographic tracing the results of these studies. However, the task of *in vivo* dissection is arduous and not suited for group studies due to its dependence on the practitioner realizing it.

This chapter sets the starting point and motivation of this thesis: the study of the human brain through *in vivo* dissection of the white matter, firstly performed by [Mori et al. \[2005\]](#) and [Catani et al. \[2002\]](#). Manually performing this dissection requires a great knowledge of brain anatomy and several hours of work. The development of a technique able to automatically perform the identification of white matter structures is, therefore, of the utmost importance. Recently, some automatic dissection algorithms which combine anatomical information with the result of diffusion MRI-based tractography have been proposed [[O'Donnell and Westin: 2007](#), [Maddah: 2008](#)]. However, these approaches need to pre-locate the bundles to be dissected, to previously build atlases of the white matter or fine-tuning parameters. This thesis addresses the automatization of *in vivo* dissection, proposing two different techniques to perform it. Firstly by directly analysing the voxels of the image obtained by diffusion magnetic resonance imaging. Secondly by clustering fibres obtained through diffusion magnetic resonance tractography. The latter clustering technique, uses a novel mathematical framework proposed in this thesis and a volumetric atlas of the white matter in order to perform a near-parameterless dissection [[Wassermann et al.: 2010a](#)]. In figure 1.14 we illustrate the efficacy of this technique by showing the re-

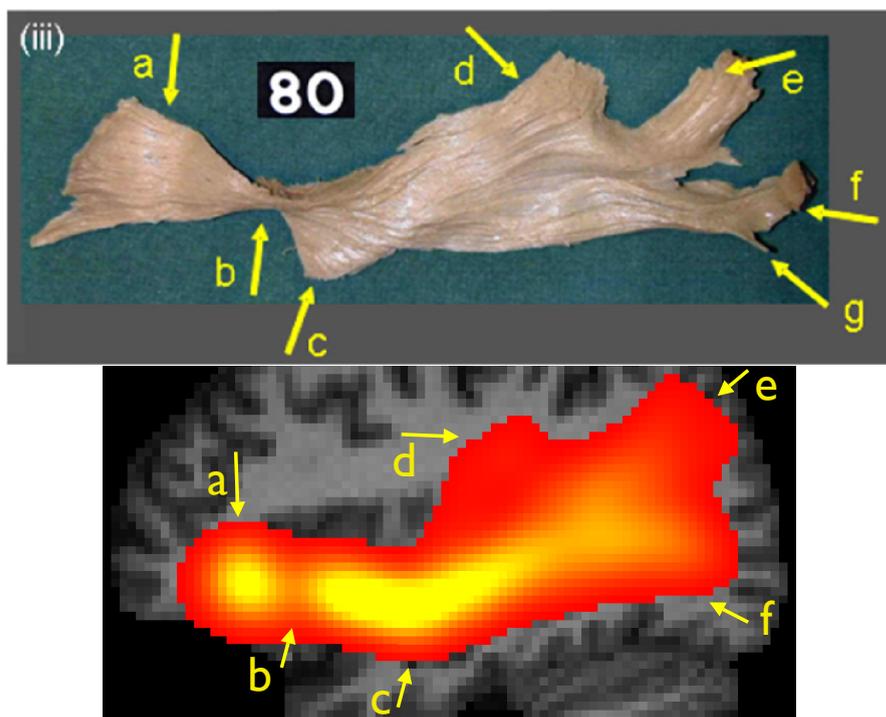


Figure 1.14: Result of the automatic dissection of the inferior-fronto occipital fasciculus using the techniques proposed in chapter 7 of this thesis. On the top, the same fasciculus excised using gross dissection [Lawes et al.: 2008], on the bottom averaged over a population of 68 individuals. The red-yellow colormap indicates the probability of the point belonging to the bundle. Red means low probability and yellow high probability.

sult of an automatic dissection of the inferior-fronto occipital fasciculus averaged over a population of 68 individuals against the same fasciculus excised using gross dissection [Lawes et al.: 2008].

1.8 SUMMARY

In this chapter we briefly went through the history of neuroscience from 3,000 BCE to the present day. Through this history we have seen how the white matter of the brain and its structure gained importance in the elucidation of brain function. Today, the advent of imaging technologies, specially diffusion magnetic resonance imaging in the last twenty years, has provided a unique means to non-invasively analyse the white matter structure, trace axonal pathologies and perform virtual dissection *in vivo*. Results obtained through this process are impressive. Still, the amount of skill and anatomical knowledge needed in order to perform it is considerable. Consequently, this thesis addresses the task of performing automatical *in vivo* dissection of the human brain white matter.

BRAIN ANATOMY AND DIFFUSION MRI

Science cannot solve the ultimate mystery of nature. And that is because, in the last analysis, we ourselves are a part of the mystery that we are trying to solve.

Max Planck (1858-1947)

We have come a long way since the Edwin Smith papyrus was written but the fundamental question remains: “How does the brain work?”. In the past 20 years, the advent of Nuclear Magnetic Resonance has allowed us to quantify *in vivo* brain function, observe its structure and correlate both approaches in order to clarify how the living brain works. Using such indirect approaches to quantify structure and function, provides rich yet complex indices of the brain. This highlights the need of mathematical models and algorithms able to build a bridge between nuclear physics and neuroscience. The main subject of this thesis is brain structure, specifically the structure of the brain’s white matter and how it can be analysed by processing the results of diffusion Magnetic Resonance Imaging. The brain being the main subject of our thesis, we start this work by briefly describing its anatomy.

2.1 HUMAN BRAIN AND NEURAL TISSUE

The most basic neural information processing unit in the brain is the neuron, a cell which transmits electrochemical signals. A neuron is composed of two main parts, the cell body which integrates information from other neurons and the axon that transmits this information, see figure 2.1a. In receiving input from other neurons, a neuron may depolarize (fire) producing an electrical action potential which travels away from the neuronal cell body, down the axon. When this action potential arrives at synapses (communicating points) with other neurons, it provokes the release of a neurotransmitter and, if there is sufficient quantity of neurotransmitter at the synaptic cleft, the receiving (postsynaptic) neuron may also fire. The conduction velocity of this potential is increased by an insulating substance, the myelin, which glial cells

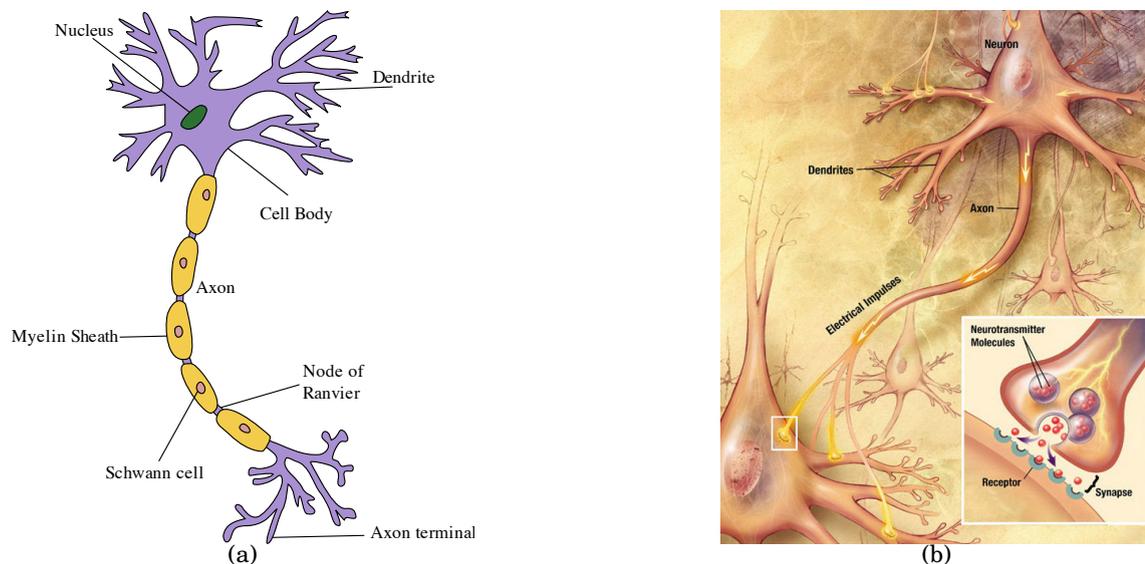


Figure 2.1: A neuron (a) and its signal propagation process (b). Adapted from Wikipedia

wrap in layers around the axon as a sheath, see figure 2.1b.

Cell bodies are gray while myelin is white. This coloring distinguishes two different components of the brain: the gray and white matter. The gray matter is principally composed of neuron bodies and its dendrites. In the human brain, the gray matter is located in the exterior part of the brain, the cortex, and in internal, or sub-cortical, brain structures like the thalamus or the basal ganglia as shown in figure 2.2. The cortex, as it can be seen in figure 2.2 as a gray layer wrapped around the brain, is a highly convoluted structure whose ridges and valleys are respectively called gyri and sulci. Millions of axons innervate the cortical and subcortical structures forming the white matter, depicted in figure 2.2 as the white or striped areas within the cortex. The white matter is composed of axonal pathways grouped into bundles which travel together called white matter fibre tracts. These tracts form a fibrous tissue as it is observable in figures 2.6 and 2.8. In addition, there are two fluid systems in the brain, the ventricles and spaces around the brain containing cortico-spinal fluid (CSF) and a vascular system providing blood flow.

The brain is divided in two hemispheres, left and right. The functions of each hemisphere are complementary, for instance, the left hemisphere controls the right hand and produces language while the right hemisphere controls the left hand and handles emotional content. Furthermore, each hemisphere is divided into five lobes, shown in figure 2.3, and the lobes into gyri, shown in figure 2.4. Each one of these lobes has a main, but not exclusive role in different aspects of brain function: the frontal lobe, responsible for motor functions, speech production, personality, insight and foresight; the parietal lobe, responsible for language comprehension, spatial orientation and perception, and somatic senses, such as touch and temperature; the temporal lobe handles audition, language comprehension, visual processing and memory;

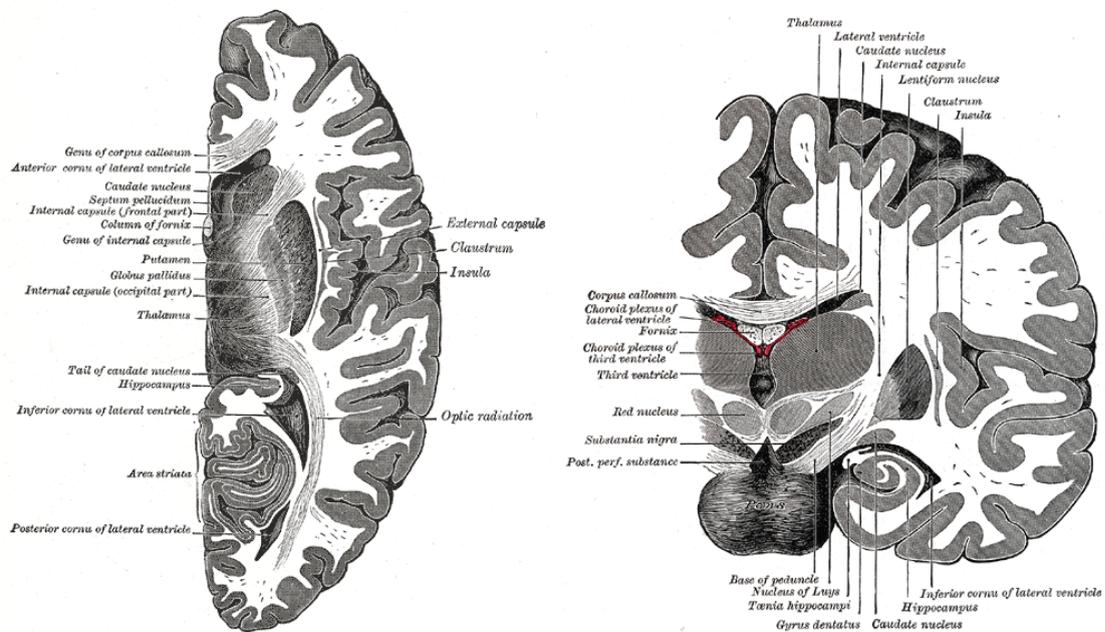


Figure 2.2: General illustrations of the brain. An axial slice on the left and a coronal slice on the right. Gray and white matter are easily distinguished. Moreover, the fibrous nature of the white matter is sketched using lines which can be observed, for instance in the optic radiation (left) and the corpus callosum (right). The vesicles and subcortical structures are marked as are all the other principal brain structures. [Gray \[1918\]](#)

the occipital lobe, responsible for primary visual processing and association; and the limbic lobe which handles emotional responses, drive-related behaviour and memory. The lobes and subcortical structures do not function in isolation, in fact they are heavily connected through the fibre bundles which compose the white matter.

White matter fibre tracts interconnect the processing streams in the functional regions of the brain. These tracts are divided in three types: commissural, connecting the hemispheres; association, connecting the lobes and gyri within the same hemisphere; and projection, connecting the spinal cord, the cerebellum and subcortical structures like the thalamus to the cortex. The fibre bundles are classified by their endpoints, like the cortico-spinal tract, or might have different names along their course like the corona radiata, internal capsule and cerebral peduncle which are named sections of the same bundle. Also, bundles are named after a region, like the cingulum or their shape, like the arcuate (arc-shaped) or uncinata (hook-shaped) fasciculi. In most cases, there are small sets of bundles which diverge from the trunk of the tract and innervate cortical and subcortical structures, as shown in figure 2.7.

2.1.1 Commissural fibres

Commissural fibres connect the right and left hemispheres, see figure 2.5. The largest and densest of these bundles is the corpus callosum (CC) composed of more than 300 million axons. This bundle connects cortical areas to their respective contralateral parts, i.e. left and right pre-central gyri, although in some cases it connects

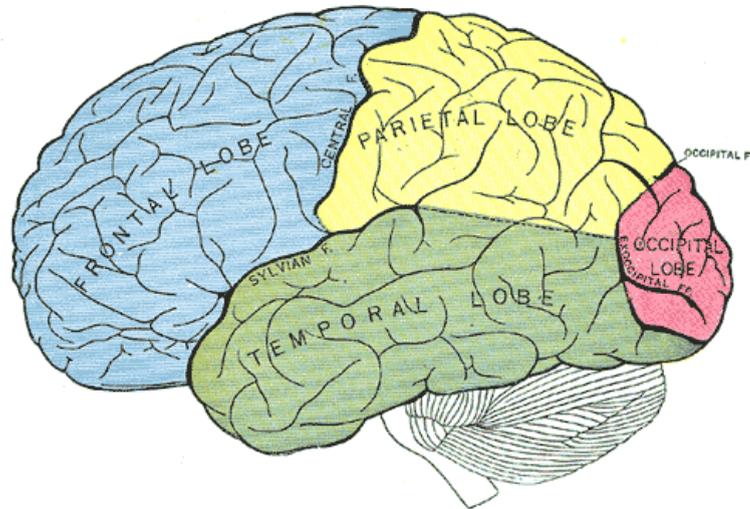


Figure 2.3: Lobes of a cerebral hemisphere. Frontal is blue, parietal yellow, temporal green and occipital red. The limbic is internal and can not be seen. Image adapted from Gray [1918]

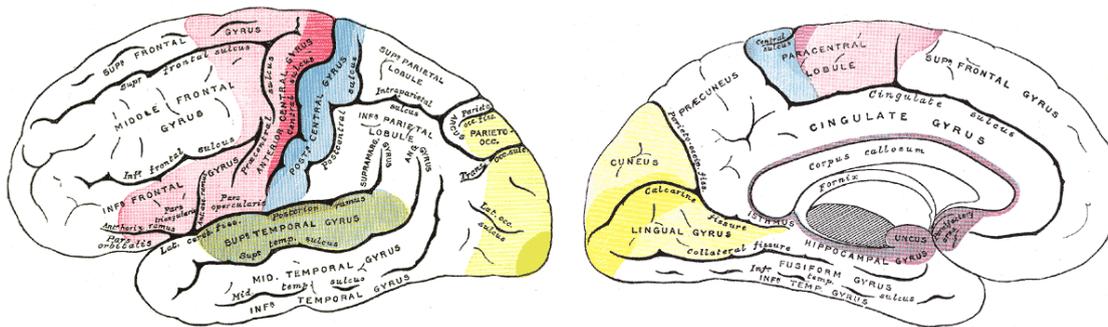


Figure 2.4: Lateral and mid-sagittal illustrations of the brain showing the most important gyri and their names. Motor area in red. Area of general sensations in blue. Visual area in yellow. Olfactory area in purple. Image adapted from Gray [1918]

related contralateral areas like the visual cortex to the contralateral visual association cortex. The anterior part of the CC is called genu, the posterior splenium and the part connecting its temporal regions is called the tapetum. The anterior (AC) and posterior (PC) commissures are smaller commissural tracts connecting the contralateral temporal cortices.

2.1.2 Association Fibres

Association fibres connect gyri within the same hemisphere. The length of their trajectories varies from short bundles connecting areas within the same gyrus and “U-fibres” connecting adjacent gyri to long associating fibres connecting the extremes of the same hemisphere as the inferior fronto-occipital fasciculus, see figure 2.6. These bundles mostly have anterior-posterior trajectories. Even if association tracts have well-defined end-points, they are composed of several smaller bundles which enter and leave the main trunk connecting cortical and sub-cortical areas, as shown in

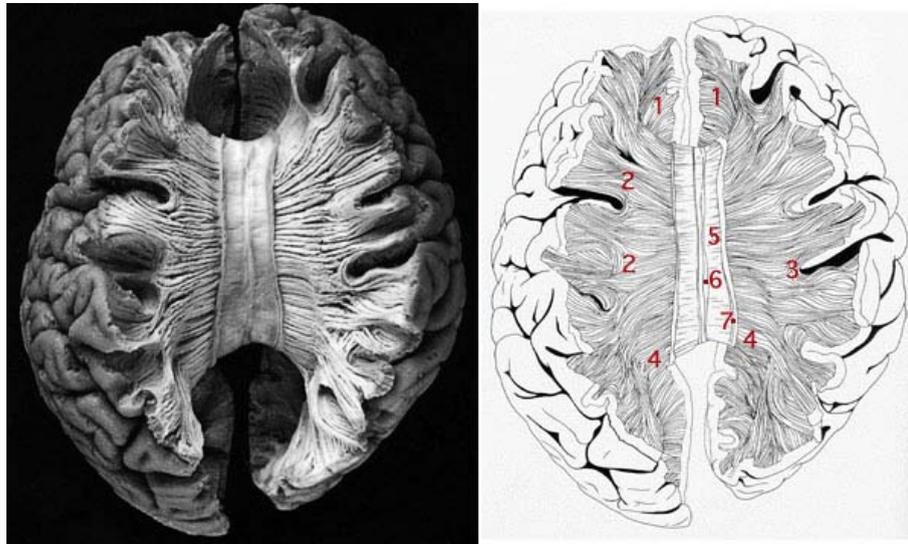


Figure 2.5: Commissural pathways. 1. Frontal forceps 2. Corpus callosum commissural fibres 3. Short arcuate fibres 4. Occipital forceps 5. Indusium griseum 6. Medial longitudinal stria 7. Lateral longitudinal stria. [Williams et al.: 1997]

figure 2.7. In this section we give an overview, which is by no means exhaustive, of the major association bundles and their function:

The cingulum (CG) and the fornix (FX), form part of the limbic system which handles memory and emotion. The cingulum (Latin word for “belt”) connects the cingulate gyrus to other regions in the limbic lobe, specially the hippocampus. The fornix goes from the hippocampus to the mammillary bodies.

The superior longitudinal fasciculus (SLF) describes an arch which connects the posterior inferior frontal cortex, specifically Brocca’s area, to Wernicke’s area in the superior posterior temporal gyrus. Traversing the inferior parietal cortex, specifically Gerschwind’s territory. This fasciculus plays a fundamental role on language and speech and it can be subdivided into three parts [Catani et al.: 2007]: the anterior indirect segment, connecting Brocca’s and Gerschwind’s areas; the posterior indirect segment connecting Gerschwind’s and Wernicke’s areas; and the arcuate describing an arch which goes through the three areas.

The uncinate fasciculus (UNC) connects the frontal portions of the temporal lobe with the inferior frontal gyrus describing a C shape which crosses behind the external capsule. Its function is still undetermined although it is believed that it forms a part of the limbic system Hasan et al. [2009].

The inferior fronto-occipital fasciculus (IFO) connects the occipital lobe to the frontal lobe running through the temporal lobe, below the insular cortex. Its role is still a subject of debate, however it is believed to play a role in spatial and visual processing [Schmahmann and Pandya: 2007b].

The inferior longitudinal fasciculus (ILF) connects the occipital and temporal

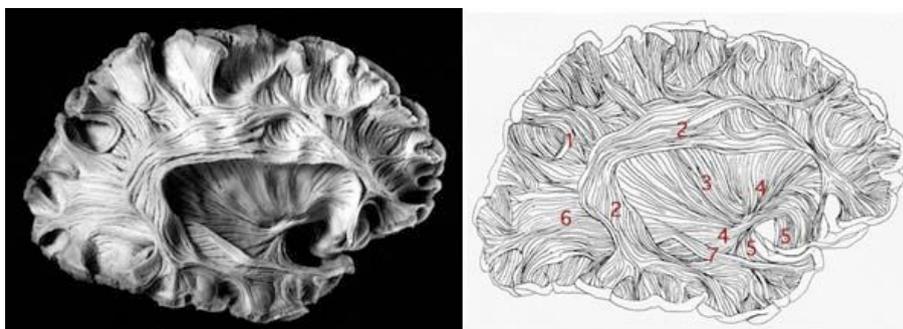


Figure 2.6: Association pathways. 1. Short arcuate bundles (U-fibres). 2. Superior longitudinal fasciculus (part is known as the arcuate fasciculus). 3. External capsule. 4. Inferior occipito-frontal fasciculus. 5. Uncinate fasciculus. 6. Sagittal stratum. 7. Inferior longitudinal fasciculus. [Williams et al.: 1997]

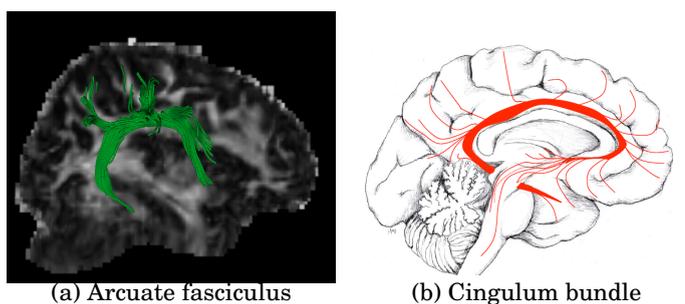


Figure 2.7: Axons enter and leave anatomical bundles. Image (a) manually selected Arcuate fasciculus after full brain tractography. Image (b), reproduced from O'Donnell [2006] with permission of Jimmy Fallon, UCI.

lobes by linking their poles. It is involved in visual processing, particularly in object recognition [Catani et al.: 2003].

2.1.3 Projection fibres

These fibres connect the cortex with the spinal cord, the cerebellum and subcortical structures like the thalamus and the basal ganglia. They compose a large bundle which starts at the pyramidal tract and goes through the cerebral peduncle, internal capsule, corona radiata and reaches the cortex, as seen in figure 2.8. This bundle is principally composed of

The cortico-thalamic radiations (CTR), connecting the thalamus with the cortex. This set of bundles has representatives which play fundamental roles in sensory tasks like the optic tract (OT) or the anterior thalamic radiation (ATR).

The cortico-pontine fibres (CP), connecting the pons with the cortex including the cortico-spinal tract (CST). This bundle is involved in sensory-motor tasks.

The cerebral peduncle (CP), connecting the cerebellum with the cortex. This set of bundles carries motor information between the brain and the body.

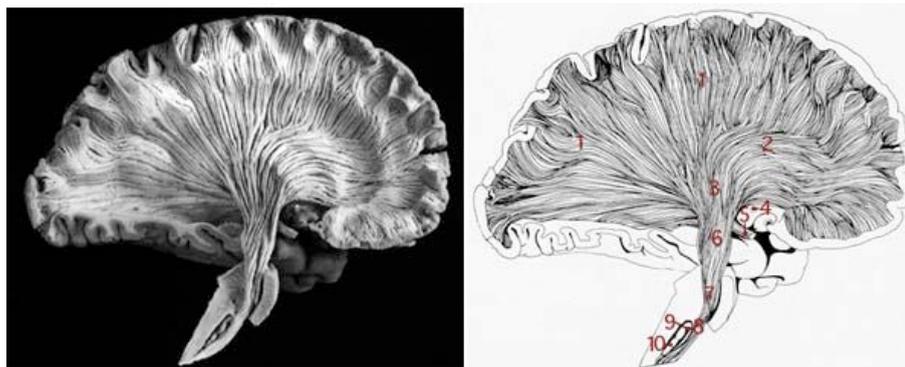


Figure 2.8: Projection fibres (mid-sagittal view). 1. Corona radiata 2. Anterior thalamic radiation 3. Internal capsule 4. Anterior commissure (which is a commissural bundle) 5. Optic tract 6. Cerebral peduncle 7. Longitudinal pontine fibres (cortico-spinal and cortico-nuclear tracts) 8. Pyramidal tract of medulla oblongata 9. Hilus of olivary nucleus 10. Olivary nucleus. [Williams et al.: 1997]

2.1.4 Microscopical structure of the white matter

The main goal of this thesis is the *in vivo* and non-invasive analysis of the human brain's white matter. The main characteristic which enables this type of study lies on the microscopical structure of the white matter and how it affects the random motion of water particles within it.

When looking closely at the white matter, as in the histological study shown in figure 2.9, we observe that it is composed of axons, or white matter fibres, packed together in bundles. Within this configuration, it is possible to divide the white matter in two compartments, an intra-axonal and extra-axonal one. Water particles in the brain are in constant motion. Particularly, those contained within the intra-axonal compartment are restricted by the axonal walls and move mostly along the axon. On the other hand, those contained in the extra-axonal compartment are able to move in any direction, even if they do not move freely as they are being hindered by the axons. The random motion of these particles, known as the physical process called diffusion, constitutes a fundamental property of the white matter tissue which, in turn, facilitates *in vivo* analysis of the white matter structure [Beaulieu: 2002].

Characterizing and quantifying the diffusion of water particles through a nuclear magnetic resonance technique called diffusion-weighted imaging, as we will see later on this thesis, enables us reconstruct the commissural, association and projective fibres which constitute this tissue and to study them. Before going into how diffusion process can be measured and how this measures can be used to characterize the structure of the white matter, we describe the diffusion process itself.

2.2 THE DIFFUSION PROCESS

Scottish botanist Robert Brown had just returned from a trip to Australia. It was 1827 and he had brought several samples of indigenous Australian flora. He devoted

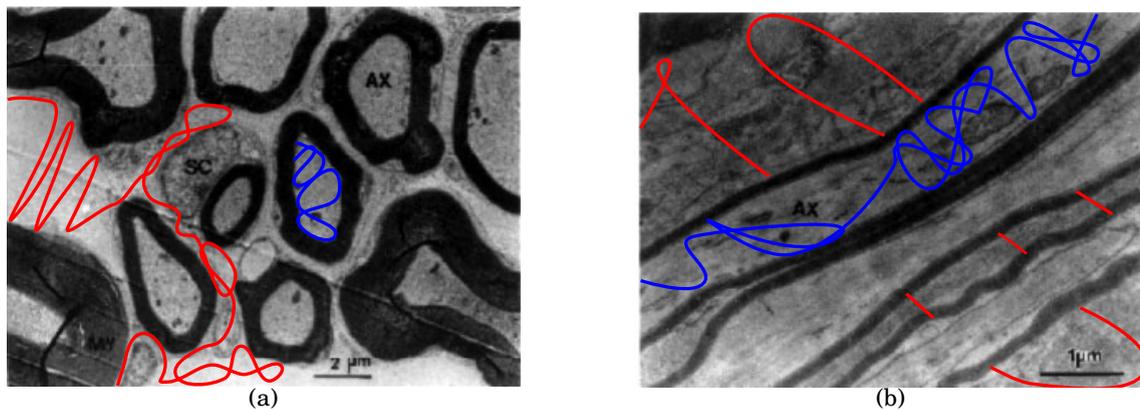


Figure 2.9: Electron micrographs of transverse (a) and longitudinal (b) sections of white matter tissue. The axons (AX), myelin sheath (MY) are labeled on the image. The random motion of a water particle in the extra-axonal compartment is illustrated as a red trajectory and the motion within the intra-axonal compartment is drawn in blue. Image adapted from [Beaulieu \[2002\]](#)

himself to the study of fertilization mechanisms and, in order to elucidate those mechanisms, he suspended a sample of pollen grains in water and observed them under his microscope. To his surprise, Brown found those grains to be in constant motion, they could even be alive. Being a man of science, he tried to prove that this movement was due to the fact that the grains were alive that he was observing “life itself”. He suspended rock and coal samples. Observing them, to his surprise, he found the same phenomenon: rock and coal grains moved as much as pollen [[Brown: 1828](#)].

Today, we know that Brown observed the constant movement of water particles under thermal agitation, a physical process known as “Brownian motion” or “diffusion”. Even if he did not observe “life itself”, the mechanism he observed is fundamental for any living being. Diffusion facilitates the transport of metabolites into cells, is omnipresent in living tissue water, and plays a fundamental role in many other vital processes of living organisms [[Jones: 2008](#)].

But, what is diffusion? Let us imagine a glass of water. If we drop a splash of dark ink in this glass, we can observe that the liquid in the glass goes through two stages:

1. Right after the ink was dropped into the water, there is a noticeable ink stain. The ink and the clear water are easily distinguishable.
2. After a while, and without the need to move the glass or stir the liquid, all the water in the glass is tinted, showing a paler tone of the original ink color.

This mixing process we have just witnessed is called “diffusion”.

If we analyse diffusion on a particle-by-particle basis we find that of the particles experienced what is called a “random walk”. By random walk, we mean that an observed particle stays for a fixed time τ in a particular place and then moves to a random new location in space. This process repeats itself without memory, meaning that the previous location has no influence on where the particle will go next. Due to

its random nature, it is not possible to predict the path of a simple particle. However, [Einstein \[1905\]](#) proved that, if we observe a sufficiently large sample of particles, it is possible to characterize at least one aspect of this phenomenon: the squared displacement of the particles from their starting point over a time τ , averaged over all the sampled particles, is directly proportional to the observation time. This is usually notated as

$$\langle r^2 \rangle = 6D\tau$$

where $\langle \cdot \rangle$ accounts for the averaging operation, r for the displacement and the scalar constant D , known as the self diffusion coefficient, measures the mobility of the particle ensemble. More precisely,

$$D = \frac{1}{6\tau} \langle \mathbf{R}^T \mathbf{R} \rangle \quad (2.1)$$

where $\mathbf{R} = \mathbf{r} - \mathbf{r}_0$ is the net displacement vector of a particle with \mathbf{r} being its position after a time τ , and \mathbf{r}_0 its original position. This model assumes that the medium where the particles are is unrestricted. The particles therefore have equal mobility in every direction. This is known as the isotropic case, where the diffusion coefficient depends only on the medium and not the direction.

Taking a step back from particle observation and looking at the diffusion process macroscopically, we find that [Fick \[1855\]](#), coauthor of what is likely the first treatise in medical physics [[Fick and Müller: 1866](#)], had previously formulated a series of laws which govern the diffusion process. Particularly, the first law of [[Fick: 1855](#)] relates the concentration difference of the diffusion substance C to a flux \mathbf{J} . This flux is proportional to the gradient of the concentration, ∇C . The proportionality constant D is the previously mentioned diffusion coefficient D and the governing equation is given by

$$\mathbf{J} = -D\nabla C. \quad (2.2)$$

One of the main concepts behind Fick's law is that particles go from areas of higher concentration to areas of lower concentration resulting in an even concentration in the whole fluid (as illustrated by the glass with ink and water), hence the minus sign. This is entirely analogous with Fourier's previous law of heat transfer, where heat flows from high to low temperature regions, and was probably an inspiration to Fick.

Focusing on the goal of this thesis, we consider the case of biological tissues. In these tissues, the mobility of a water molecule is constrained by obstacles formed by surrounding structures. In chapter 2, particularly in figure 2.9, we have seen that this is the case of white matter tissue in the brain [[Beaulieu: 2002](#)]. The general diffusion displacement probability density function (p.d.f.), also called diffusion propagator, of water molecules is extremely complex and is still unknown today. Therefore, simple models of diffusion have been historically proposed. Amongst these models, the most popular is certainly the diffusion tensor (DT) model proposed

by [Basser et al. \[1994b\]](#). In this simplified model of water diffusion, Einstein's and Fick's laws of diffusion are generalized. The scalar diffusion coefficient D is replaced by a positive symmetric semi-definite matrix \mathcal{D} representing diffusion, the diffusion tensor. Therefore, Einstein's relation (equation 2.1) is generalized, considering the covariance matrix of the net displacement vector \mathbf{R}

$$\mathcal{D} = \begin{pmatrix} \mathcal{D}_{xx} & \mathcal{D}_{xy} & \mathcal{D}_{xz} \\ \mathcal{D}_{xy} & \mathcal{D}_{yy} & \mathcal{D}_{yz} \\ \mathcal{D}_{xz} & \mathcal{D}_{yz} & \mathcal{D}_{zz} \end{pmatrix} = \frac{1}{6\tau} \langle \mathbf{R}\mathbf{R}^T \rangle. \quad (2.3)$$

We can use this generalization to characterize the diffusion propagator. First, we notate the probability that a particle moves along the vector \mathbf{R} in a time τ as $P(\mathbf{R}, \tau)$. Then, we make a first order approximation to $P(\mathbf{R}, \tau)$ ignoring the high order terms and using equation 2.3. Finally, we obtain a partial differential equation which governs the diffusion propagator¹:

$$\frac{\partial P(\mathbf{R}, \tau)}{\partial \tau} = \mathcal{D} \nabla^2 P(\mathbf{R}, \tau). \quad (2.4)$$

Under the assumption that the diffusion is Gaussian, the solution of this equation is the propagator model given by [Basser et al. \[1994b\]](#):

$$P(\mathbf{R}, \tau) = \frac{1}{\sqrt{(4\pi\tau)^3 |\mathcal{D}|}} \exp\left(-\frac{\mathbf{R}^T \mathcal{D}^{-1} \mathbf{R}}{4\tau}\right).$$

The characteristics of this propagator model have proved to be an invaluable resource in characterizing diffusion in living tissue, particularly in the human brain [[Johansen-Berg and Behrens: 2009](#)].

In summary, diffusion is a fundamental physical process in nature and particularly physiology. The random motion of water molecules within a tissue are influenced by a variety of factors like cell membranes, the cytoskeleton, and macromolecules [[Tanner and Stejskal: 1968](#)]. It is due to this that being able to measure and characterize diffusion, and in fact the diffusion propagator, is a valuable tool to elucidate the microstructural and physiological features of tissues [[Basser and Pierpaoli: 1996](#)]. In the past twenty years, nuclear magnetic resonance has made a breakthrough in this area providing, through diffusion magnetic resonance imaging (dMRI), a probe into the microstructure of living tissue [[Alexander et al.: 2007](#)].

2.3 NUCLEAR MAGNETIC RESONANCE

In 1946, [Bloch et al. \[1946\]](#) and [Purcell et al. \[1946\]](#) simultaneously described Nuclear Magnetic Resonance (NMR) which yielded them a joint Nobel Prize in Physics in 1952. The basic principle behind NMR is that after aligning a magnetic nu-

1. The detailed intermediate mathematical derivations can be found in the work by [Tuch \[2002\]](#)

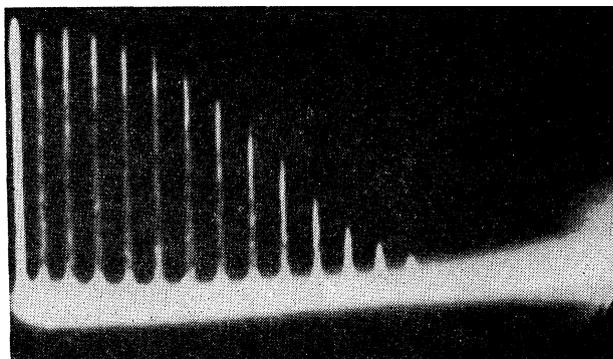


Figure 2.10: Water diffusion measured by the first NMR-obtained signal: The decay of transverse polarization associated with water at 25°C used by Carr and Purcell [1954] to measure the diffusion coefficient of water. In this work, using a small magnetic gradient and the Spin-Echo method proposed earlier by Hahn [1950] they observed that the decay is largely determined by the molecular diffusion through the magnetic gradient. Image adapted from Carr and Purcell [1954]

cleus, for instance the proton ^1H , with a very strong external magnetic field, its response to a perturbation of the alignment by an electromagnetic field is characteristic. Shortly after this, Hahn [1950] published a paper on the NMR spin echo in which he noted that the random thermal motion of the spins would reduce the amplitude of the observed signal in the presence of an inhomogeneity of the magnetic field. His work was influential on NMR and fundamental in understanding magnetic resonance imaging (MRI). Inspired by this work, Carr and Purcell [1954] used the Spin-Echo NMR sequence proposed by Hahn [1950] and a sequence of their own to measure the diffusion coefficient of water at 25°C , see figure 2.10. Carr and Purcell used a small magnetic gradient and the Spin-Echo sequence and showed that the decay on the transverse polarization was affected by water diffusion. Then they proposed a second NMR sequence which was not influenced by water diffusion and used both decays obtained by both sequences to measure the diffusion constant. This was the first NMR signal ever obtained.

Almost 20 years later, the first acquisition of a bi-dimensional image using NMR was performed by Lauterbur [1973], see figure 2.11. As the image was obtained using two coupled magnetic gradients, he called the process zeugmatography, from the Greek work zeugma, “that which is used to join”. Mansfield [1977] improved on the previous technique. By using mathematical properties of the MRI signal, he proposed a new ultrafast acquisition procedure: echo-planar imaging. The two previous techniques, zeugmatography and echo-planar imaging, became a fundamental parts of medical MRI yielded a joint Nobel Prize in Physiology or Medicine 2003 to their authors. There is a single piece missing to fully describe the techniques at the heart of diffusion MRI: the gradient spin echo sequence developed by Stejskal and Tanner [1965]. Due to its critical role, we dedicate the next section to the description of their technique.

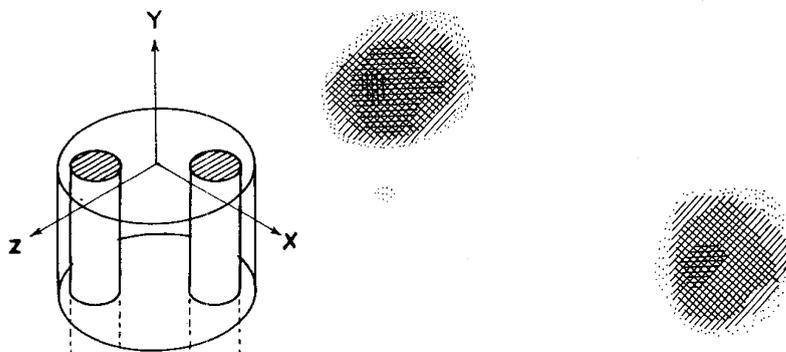


Figure 2.11: First bi-dimensional image obtained from NMR. [Lauterbur \[1973\]](#) was the first to propose the use of two coupled magnetic gradients in order to obtain a bi-dimensional slice. He called this process zeugmatography. On the left we can see a drawing of the imaged object, with two capillaries filled with water, and on the right the resulting image obtained by NMR. Both images adapted from [Lauterbur \[1973\]](#).

2.3.1 Pulse Gradient Spin Echo (PGSE)

The [Stejskal and Tanner \[1965\]](#) imaging sequence is used to measure the diffusion of water molecules in a given direction \mathbf{g} . This sequence uses two gradient pulses, $\mathbf{g}(t)$, of duration time δ , in the direction \mathbf{g} to control the diffusion-weighting. The gradient pulses are placed before and after a 180° refocusing pulse (RF). More specifically, a first 90° RF is applied to flip the magnetization in the transverse plane. The first gradient pulse induces a phase shift of the spins whose position are now a function of time. The position of the spins is assumed to stay constant during time δ . Finally, after a time Δ , the 180° combined with the second gradient pulse causes a second phase shift. For static spins only, this pulse cancels the phase shift. Alternatively, spins under Brownian motion during the time period Δ , undergo different phase shifts by the two gradient pulses, resulting in a T2 signal attenuation [[Cercignani and Horsfield: 2001](#)]. This pulse sequence is illustrated in figure 2.12

Figure 2.13 shows examples of diffusion weighted images of the brain acquired at different directions. In this figure it can be observed that direction specific attenuation is related to the orientation of white matter fibres. By assuming that the pulses are infinitely narrow (narrow pulse approximation), meaning the gradient pulse duration δ is short enough to neglect the diffusion of the water molecule at that time, [Stejskal and Tanner \[1965\]](#) showed that the attenuation of the signal $S(\mathbf{q}, \tau)$ is expressed as the 3-dimensional Fourier transform \mathcal{F} of the ensemble average propagator P ,

$$\frac{S(\mathbf{q}, \tau)}{S_0} = \int_{\mathbb{R}^3} P(\mathbf{r}|\mathbf{r}_0, \tau) \exp(-2\pi i \mathbf{q}^T \mathbf{R}) d\mathbf{r} = \mathcal{F}[P(\mathbf{r}|\mathbf{r}_0, \tau)]. \quad (2.5)$$

In this equation, $\mathbf{q} = \gamma \delta \mathbf{G} / 2\pi$, with γ the nuclear gyromagnetic ratio for water protons, \mathbf{G} the applied diffusion gradient vector, S_0 is the baseline image acquired without any diffusion gradient (also called the B_0 image) and $P(\mathbf{r}|\mathbf{r}_0, \tau)$ is the diffusion p.d.f. or diffusion propagator introduced in chapter 2. This $P(\cdot)$ is ultimately the

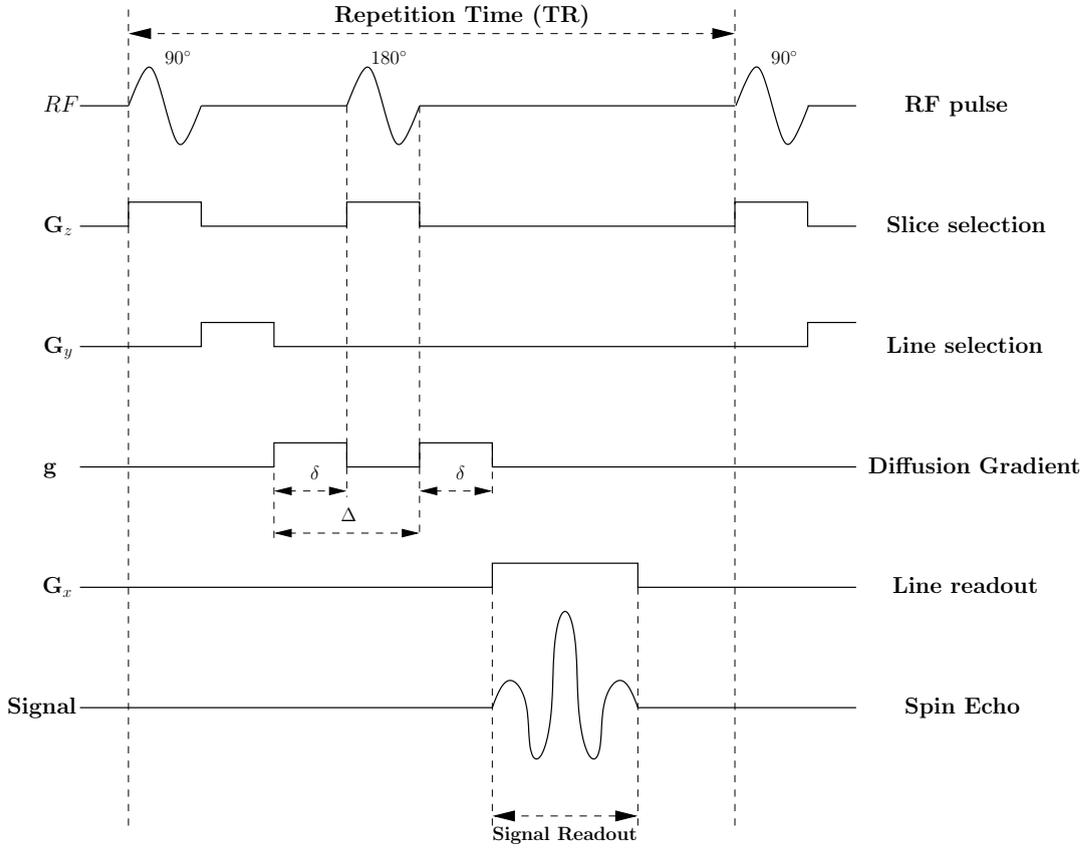


Figure 2.12: Scheme of Stejskal-Tanner imaging sequence [Stejskal and Tanner: 1965] reproduced from Lenglet [2006]

function we are seeking to reconstruct in diffusion MRI. In order to reconstruct this p.d.f. , it must be sampled along many \mathbf{q} vectors. The space of all possible 3D \mathbf{q} vectors is called \mathbf{q} -space. Callaghan et al. [1988] proposed the use of this sampling technique to reconstruct the diffusion propagator and called it \mathbf{q} -space imaging.

If we assume the diffusion p.d.f. to be Gaussian, the Fourier integral in equation 2.5 can be worked out analytically. The Stejskal-Tanner signal attenuation equation then becomes

$$\frac{S(\mathbf{q}, \tau)}{S_0} = \exp(-\tau \mathbf{q}^T \mathcal{D} \mathbf{q}) = \exp(-\tau D(\mathbf{q})), \quad (2.6)$$

where $D(\mathbf{q}) = \mathbf{q}^T \mathcal{D} \mathbf{q}$ is the apparent diffusion coefficient (ADC). The signal attenuation is often written using a unit vector $\mathbf{g} = \mathbf{q}/\|\mathbf{q}\|$ and a quantity called b-value [Le Bihan et al.: 1986], $b = \tau \|\mathbf{q}\|^2$:

$$\frac{S(b, \mathbf{g})}{S_0} = \exp(-b \mathbf{g}^T \mathcal{D} \mathbf{g}) = \exp(-b D(\mathbf{g})), \text{ where } S_0 = S(0, \cdot). \quad (2.7)$$

This is the most common formulation under a Gaussian assumption. For large b-values, the true signal quickly falls off, while the background noise is relatively unaf-

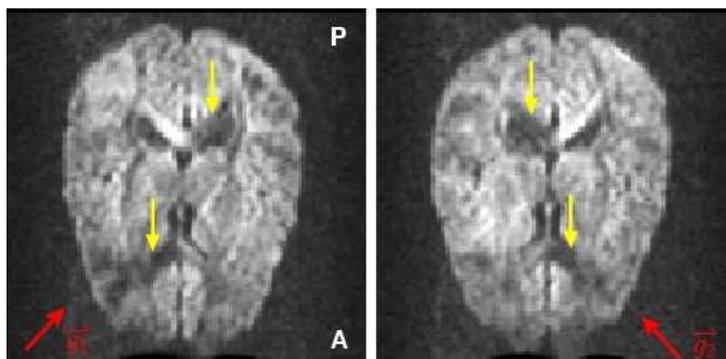


Figure 2.13: Axial slice of diffusion-weighted images (DWI) with two different diffusion gradient directions (red arrows). MR signal attenuation is found in regions having fibres mostly aligned with diffusion gradient direction (yellow arrows).

ected resulting in noisy measurements, shown in figure 2.14. The b -value has to be appropriately tuned to avoid either very low signal attenuation when b is low or poor signal-to-noise ratio (SNR) when b is high.

2.4 DIFFUSION-WEIGHTED MRI OF LIVING TISSUE

The first diffusion-weighted Magnetic Resonance Imaging (DWI) acquisition of biological tissue took place in 1984. [Websey, Moseley, and Ehman](#) [[Websey et al.: 1984a,b](#)] obtained an image of *in vitro* biological tissue followed by [Taylor and Bushell](#) [[1985](#)] who imaged a hen's egg. However, none of these acquisition techniques were suitable for use in a clinical setting. Less than a year after this, [Le Bihan and Breton](#) [[Le Bihan and Breton: 1985](#), [Le Bihan et al.: 1986](#)] were the first who performed a DWI acquisition *in vivo* of the human brain using a whole-body scanner. Up to this point, a DW image was simply the unprocessed result of the application of Stejskal and Tanner's pulse sequence in one gradient direction, which corresponded to a single point in q -space [[Callaghan et al.: 1988](#)].

2.4.1 Anisotropic diffusion in living tissue

[Moseley et al. \[1990\]](#) were the first to use various gradient directions in order to show heavy DW anisotropy in living tissue. They proposed the apparent diffusion coefficient (ADC) to quantify diffusion in a given direction. By assuming that diffusion within an imaged voxel followed a Gaussian p.d.f. and using the Gaussian-based solution of the Stejskal-Tanner equation 2.7, they calculated the ADC as

$$D(\mathbf{g}) = -\frac{1}{b} \ln \left(\frac{S(b, \mathbf{g})}{S_0} \right).$$

Through this formulation they argued that signal attenuation in a PGSE sequence measured apparent diffusivity parallel to the sensitizing gradient. Their first experiment showing that diffusion was anisotropic used a segment of fresh sugar cane.

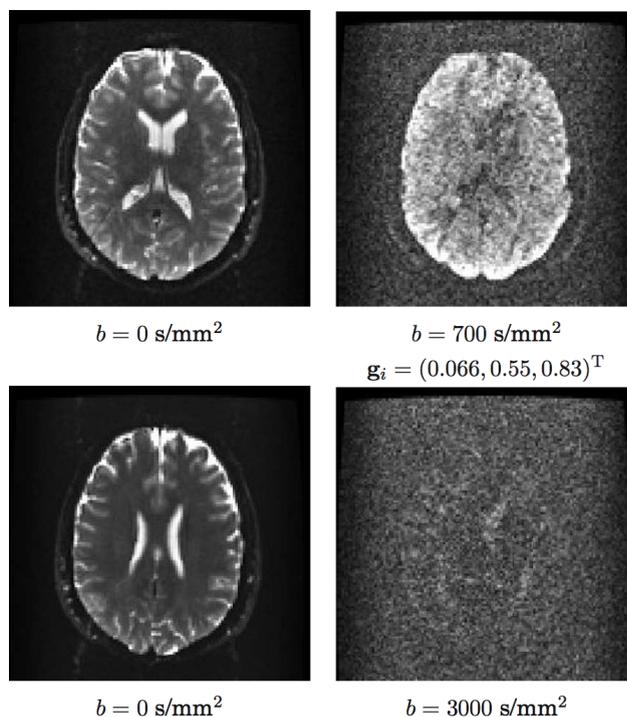


Figure 2.14: Samples of DW images with different b-values. Its noticeable the decay in the signal-to-noise ratio as the b-value increases.

They imaged two directions perpendicular to the circular section of the sugar cane and found that images showed differences in the areas where its vascular fibres of the cane were aligned with the gradients. Moreover, they showed that in different white matter structures of a cat, signal on the x (right-left), y (rostral-caudal), and z (dorsal-ventral) axes was attenuated when the fibres of the structures were parallel to the diffusion-sensitizing gradient.

The attenuation of the signal (increase of the ADC) observed by [Moseley et al.](#) in the cat's brain was due to the restricted nature of diffusion inside white matter fibres or axons. Myelin, the axonal membrane, microtubules and neurofilaments are all longitudinally oriented structures that could hinder water diffusion perpendicular to the length of the axon and cause the perpendicular diffusion coefficient $D(\perp)$ to be smaller than the parallel diffusion coefficient $D(\parallel)$ [[Beaulieu: 2002](#)] as illustrated in figure 2.15a. When the apparent diffusion coefficients are measured in a voxel, as illustrated in figure 2.15b, there is diffusion taking place inside (blue) and outside the axons (red). However, the diffusion outside the axons is negligible. The signal attenuation therefore corresponds to the diffusion within them.

Up to this point, directional differences in diffusion had been shown in an organized tissue, the brain's white matter. The goal was now to find a formalism that, unlike the ADC, would not be as dependant of the gradient encoding used for the acquisition. It was [Basser et al.](#) [[Basser et al.: 1994a,b](#)] who, in 1992, proposed the use of a symmetric and positive-definite matrix to model the intrinsic diffusion properties

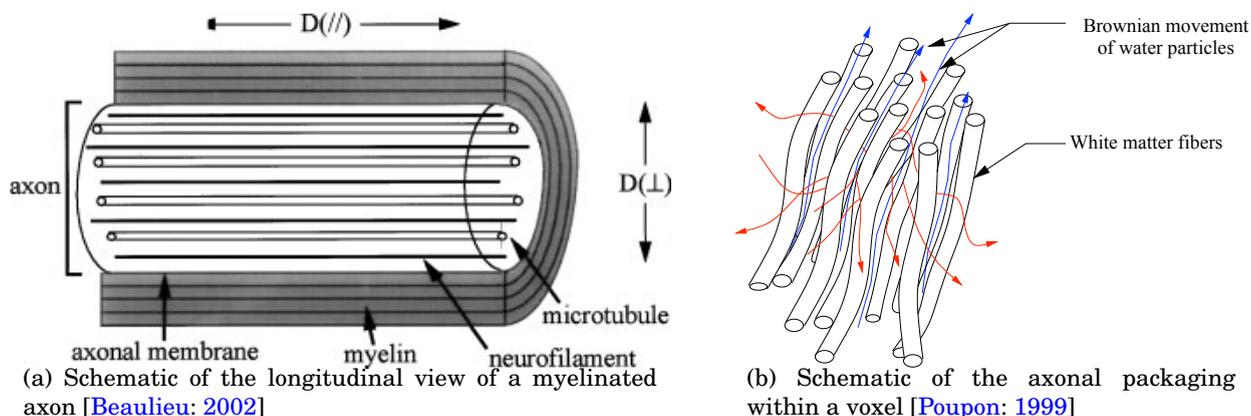


Figure 2.15: The restricted nature of diffusion inside white matter fibres or axons: Myelin, the axonal membrane, microtubules and neurofilaments are all longitudinally oriented structures that could hinder water diffusion perpendicular to the length of the axon and cause the perpendicular diffusion coefficient $D(\perp)$ to be smaller than the parallel diffusion coefficient $D(\parallel)$ [Beaulieu: 2002] as is illustrated in (a). When the apparent diffusion coefficients are measured in a voxel, as illustrated in (b), there is diffusion taking place inside (blue) and outside the axons (red). However, the diffusion taking place outside of the axons is negligible, hence the signal attenuation corresponds to the diffusion inside of the axons.

of biological tissues, a second order diffusion tensor (DT). They called this particular type of DWI diffusion tensor magnetic resonance imaging (DTI).

2.4.2 Diffusion Tensor MRI

In introducing DTI, Basser et al. made two fundamental points with regards to previous research [Basser et al.: 1994a]:

1. When imaging a tissue in which the diffusion was anisotropic, it was not always possible to know the preferred diffusion direction. Therefore, using explicit-chosen gradient-sensitizing directions, as Moseley et al. [1990] did, was not always convenient.
2. Measuring diffusion only along the three coordinate axes, the diagonal of the DT was not sufficient to fully characterize anisotropic diffusion. All the elements of the DT should be calculated, diagonal and off-diagonal ones.

To overcome these limitations, they assumed that the diffusion propagator had a Gaussian p.d.f. and proposed to use the full DT to model diffusion in each imaged voxel. Using the Gaussian p.d.f. assumption, they solved the differential equation 2.4. This equation characterises the diffusion propagator, and established the DT model for the diffusion propagator:

$$P(\mathbf{r}|\mathbf{r}_0, \tau) = \frac{1}{\sqrt{(4\pi\tau)^3 |\mathcal{D}|}} \exp\left(-\frac{(\mathbf{r} - \mathbf{r}_0)^T \mathcal{D}^{-1} (\mathbf{r} - \mathbf{r}_0)}{4\tau}\right),$$

which accounts for the probability that a particle moves from \mathbf{r}_0 to \mathbf{r} in a time τ . In fact, the DT model approximates the propagator by a 3-variate normal distribution

with zero mean. The diffusion tensor \mathcal{D} therefore can be viewed as the covariance matrix describing the Brownian motion of an ensemble of water molecules, shown in equation 2.3, in each imaging voxel. Having proposed a model for the propagator which provided a solution for the previously mentioned issues, the next problem to tackle was how to estimate \mathcal{D} from a set of DW images.

Since \mathcal{D} is symmetric (equation 2.3), it has only six unknown coefficients that we need to estimate. Therefore, DTI needs at least six DW images and one unweighted diffusion image ($b = 0s/mm^2$), usually called B_0 , to solve the system of equations which has the coefficients of the DT as a solution. The typical setting consists of between 7 and 60 DW images acquired with non-collinear gradient directions and $b = 1000s/mm^2$ and a single B_0 image. A large group of techniques has been proposed to estimate the DT from a set of DW images ranging from linear and non-linear least squares [Basser et al.: 1994a] to much more sophisticated techniques based on Riemannian frameworks [Lenglet et al.: 2006b, Fillard et al.: 2007].

Perhaps the most important feature of the DT model is the combination between its richness as a mathematical model and the simplicity of its manipulation. The DT a positive semi-definite matrix, therefore it can be decomposed in a set of positive eigenvalues, $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and its corresponding unit-norm eigenvectors e_1, e_2, e_3 . The first vector gives the principal direction of the DT, the other two span an plane orthogonal to it and the eigenvalues quantify the diffusivity in those directions. Moreover, if $\lambda_1 \gg \lambda_2$, then e_1 is aligned with the preferred diffusion direction of the ensemble of water particles inside the voxel and λ_1 is its diffusivity. As a direct consequence of the previous property, when the imaged voxel contains a bundle of well organized axons sharing a common orientation, as in figure 2.15b, the first eigenvalue is aligned with the direction of the axons.

With respect to visualization, given that the DT is a covariance matrix, it is possible, and common, to represent it as an ellipsoid [Basser et al.: 1994b, Basser: 1995], shown in figure 2.16a. Such visualization provides a sense of the anisotropy (shape) and the diffusivity (volume) of the tensor as observed in figure 2.16b. This figure shows the tensor field corresponding to an axial slice of the human brain. Tensors imaged on the CC, for instance, look very elongated, while tensors imaged on the ventricles look bigger and spherical. To achieve a simple visualization of the white matter structures, Pajevic and Pierpaoli [1999] proposed a DT-based colour scheme of the human white matter: they coloured the voxels according to the direction of the first eigenvector of the DT for tensors with a preferred direction aligned with the left-right axis; red, green for anterior-posterior; and blue for superior-inferior. In proposing this map, they showed that several white matter fibre bundles were easily identifiable due to their defined orientation, as can be seen in figure 2.16c.

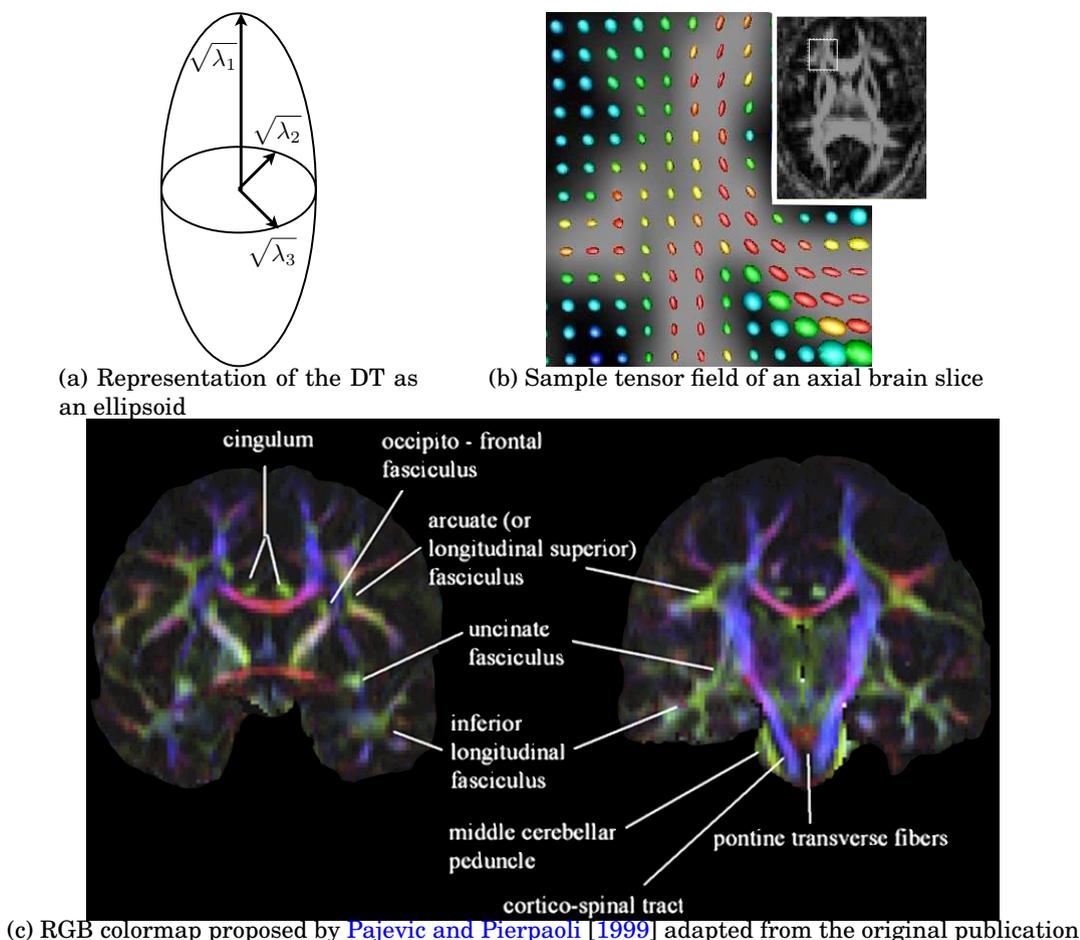


Figure 2.16: Representation and visualization of DTI: (a) The DT is a covariance matrix, and can therefore be visualized as an ellipsoid with its principal axes inferred from the DT eigendecomposition. (b) The ellipsoid representation provides a sense of the anisotropy (shape) and diffusivity (size) of the DT. (c) Using the principal diffusivity direction a color visualization which renders possible to identify major white matter structures.

DTI measures

The DT is a rich formalism providing a considerable amount of information about the diffusion within each voxel. However, manipulating matrices, especially when looking for statistical differences between populations can be cumbersome. Shortly after defining the DTI, Basser tackled this problem by defining a set of scalar, rotationally invariant measures to quantify several characteristics of the DT [[Basser: 1995](#)]. In the following we only describe the most frequently used measures:

- The trace of the diffusion tensor,

$$\text{Trace}(\mathcal{D}) = \lambda_1 + \lambda_2 + \lambda_3,$$

is a rotationally invariant index of the overall diffusivity within the voxel being 3 times the mean diffusivity, $\langle \lambda \rangle = (\lambda_1 + \lambda_2 + \lambda_3) / 3$, of the particle ensemble

contained by the voxel.

- The fractional anisotropy (FA)

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

is perhaps the most widely used index in clinical applications [Ciccarelli et al.: 2008]. It is a rotationally invariant, dimensionless measure of the anisotropy of the tensor. It ranges from 0, when the tensor represents completely isotropic diffusion (is sphere-shaped), to 1, when its diffusivity is constrained along a single axis.

- The relative anisotropy (RA)

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

corresponds to the mean deviation of the diffusivities and represents the ratio of anisotropy over the mean diffusivity of the tensors.

- Axial (λ_{\parallel}) and radial diffusivities (λ_{\perp}),

$$\begin{aligned}\lambda_{\parallel} &= \lambda_1 \\ \lambda_{\perp} &= \frac{\lambda_2 + \lambda_3}{2},\end{aligned}$$

provide a sense of the quantity of diffusion taking place along the preferred diffusion direction and orthogonal to it. Histological studies using mice and other phantoms report that these two measures used together provide an indicator demyelination and dysmyelination of the axons [Song et al.: 2002, Beaulieu: 2002]. Due to their physiological interpretation, these measures are frequently used in order to assess white matter integrity in neurological disorders [Ciccarelli et al.: 2008].

- On a more geometrical note, linear (C_L), planar (C_P) and spherical (C_S) measures

$$\begin{aligned}C_L &= \frac{\lambda_1 - \lambda_2}{\lambda_1} \\ C_P &= \frac{\lambda_2 - \lambda_3}{\lambda_1} \\ C_S &= \frac{\lambda_3}{\lambda_1}\end{aligned}$$

proposed by Westin et al. [2002], describe whether the tensor is cigar-shaped (C_L), disk-shaped (C_P) or spherical-shaped (C_S).

Although these measures have been successfully used in different applications, all but radial and axial diffusivity share a common limitation: the measures are cou-

pled. This means that when one of the measures varies, the others variate too. Hence, looking only at variations on these measures it is not simple to specify in what way the tensor properties were affected. In order to solve this, [Ennis and Kindlmann \[2006\]](#) proposed two sets of three measures each to quantify, orthogonally, shape and diffusivity of the tensors. On the other hand, [Savadjiev et al. \[2010\]](#) calculate scalar measures on neighbourhoods of voxels in order to describe local structure of the white matter in an area and not in a single voxel. Even if these recent measures have not been tested extensively, studies are currently underway applying them to clinical settings in the quest for biomarkers.

2.4.3 Tensor computation

The first question arising tensor computation concerns the validity of the tensors: Is every symmetric 3×3 positive definite matrix a valid DT? The answer is no, in the physical setting which tensors model there are no possible scenarios where an eigenvalue might be 0 [\[Fillard et al.: 2007\]](#). This is due to the nature of the diffusion process: even if the diffusion is clearly oriented in one direction and very restrictive in the perpendicular plane, the case where a biological tissue might not let a water particle move in a certain direction is impossible. Consequently, the DT constitute a convex set, in particular isometric to the shell of a cone in \mathbb{R}^6 , where no 0 eigenvalues are permitted [\[Fillard et al.: 2007, Arsigny et al.: 2006, Lenglet et al.: 2006b\]](#).

From the application perspective, several operations on DTI like denoising or estimation of the tensor values rely heavily in complex computations like solving partial differential equations or gradient descent [\[Fillard et al.: 2007, Castaño-Moraga et al.: 2007\]](#). There are several difficulties in these computations on the space of DTs. Due to the fact that while in standard Euclidean calculus convex operations are stable, for instance: mean of a set of tensors is a tensor, the boundaries of the space can easily be reached with the complex computations we mentioned before and null or negative eigenvalues of the DT might appear.

Even if very small eigenvalues are clearly not null ones, both are unlikely to exist from a physical point of view. Consequently, a mathematical framework suited for DT computation must render improbable that such tensors result from computation. In order to achieve this, several mathematical frameworks have been developed among all of them, we only introduce the most used two: the Riemannian [\[Lenglet et al.: 2006b\]](#) and the Log-Euclidean [\[Arsigny et al.: 2006\]](#) frameworks.

[Lenglet et al. \[2006b\]](#) and [Arsigny et al. \[2006\]](#) proposed the Riemannian and Log-Euclidean frameworks respectively with the two goals: the operations between tensors should stay in the convex cone that we described before; and given two tensors, a gradual interpolation between them must preserve monotone interpolation of the determinants. These two frameworks are equivalent with a distinction on their implementation. As all of the properties and operations of these frameworks are defined after the metric, we only introduce the metric proposed by them:

The metric proposed by [Lenglet et al. \[2006b\]](#) uses the idea that the tensors are the parameter of a zero-mean Gaussian distribution and use Fisher’s information matrix in order to derive it. Finally the formulation their metric, the Geodesic metric between tensors, is

$$d_g(\mathcal{D}_1, \mathcal{D}_2) = \sqrt{\frac{1}{2} \sum_{i=1}^3 \ln^2(\lambda_i)} \quad (2.8)$$

Where the scalars λ_i are the eigenvalues of the matrix $\mathcal{D}_1^{-\frac{1}{2}} \mathcal{D}_2 \mathcal{D}_1^{-\frac{1}{2}}$.

Similarly, [Arsigny et al. \[2006\]](#) proposed the Log-Euclidean metric. This metric which shares the same interpretation than the previous one, based on Lie algebras, uses matrix-exponentials as an alternative to calculate it:

$$d_{le}(\mathcal{D}_1, \mathcal{D}_2) = \sqrt{\text{trace}(\{\log(\mathcal{D}_1) - \log(\mathcal{D}_2)\}^2)} \quad (2.9)$$

where in this case $\log(\cdot)$ refers to the matrix logarithm.

Both of these metrics yield close results and have been used in clinical and image-processing applications, however the Log-Euclidean version has the advantage of providing a faster environment to perform tensor calculations.

Several other metrics have been proposed, among them the J-Divergence [[Wang and Vemuri: 2005](#)] based on the Kullback-Leibler divergence, the Geodesic-Loxodromes [[Kindlmann et al.: 2007](#)] which aim to maintain FA monotone in interpolation operations, and many others.

2.4.4 Beyond DT-MRI

There is a considerable amount of (constantly growing) work using the DT diffusion model and its derivatives in clinical neurology and in human brain function research [[Jones: 2008](#), [Ciccarelli et al.: 2008](#), [Kubicki et al.: 2007](#)]. However, [Wiegell et al. \[2000\]](#) proved that at the scale to which an MRI machine is capable of performing imaging, DTI has shortcomings in representing water diffusion in the brain’s white matter. While the radius of an axon varies from $0.2\mu m$ to $20\mu m$, the volume of a DW voxel ranges from $1mm^3$ to $3mm^3$ in clinical settings. Consequently, when different axon bundles within a voxel cross, kiss or diverge, several directions could be regarded as the “preferred diffusivity”. In these cases, in which a single fibre orientation is completely inadequate in characterizing the fibre orientations, the DT tensor formalism, which assumes an unimodal p.d.f., is insufficient. This highlights the need for better diffusion models, capable of allowing more than a single preferred fibre orientation.

A solution to this problem, QBall Magnetic Resonance Imaging, was proposed by [Tuch \[2004\]](#). This solution was improved later by [Descoteaux et al. \[2007a\]](#) among others. In the following section we go through the details of the QBall Magnetic Resonance Imaging (QBI) technique presented by [Descoteaux et al. \[2007a\]](#) which we

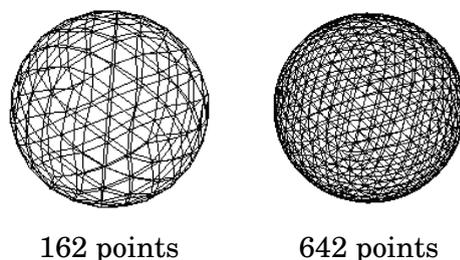


Figure 2.17: Discrete samplings of the sphere for different numbers of gradient directions corresponding to order 3 and order 4 tessellation of the sphere respectively.

use in this thesis.

2.4.5 QBall MRI

As mentioned before, the DT model lacks the capability of accurately modeling the diffusion of water particles where there are several fibres crossing the voxel in different directions. In order to overcome this limitation, [Tuch \[2004\]](#) proposed QBall Magnetic Resonance Imaging (QBI). In this method, the angular information of the diffusion probability displacement of water molecules is modelled through a diffusion Orientation Distribution Function (ODF). In order to obtain a probability of diffusion for the particle ensemble in each direction, the ODF is reconstructed as by integrating the propagator along its radius,

$$\Psi(\mathbf{r}) = \frac{1}{Z} \int_0^\infty \mathcal{P}'(\zeta \mathbf{r}) d\zeta, \mathbf{r} \in S^2, \quad (2.10)$$

where S^2 is the unit-norm sphere, Z is a dimensionless normalization constant and $\mathcal{P}'(\cdot)$ is the diffusion propagator. Consequently, the ODF is an angular function in which, for each direction, it expresses the diffusion probability. Recently, the formulation of the previous equation has been corrected by [\[Aganj et al.: 2009\]](#) and [Tristán-Vega et al. \[2009\]](#) who added the missing volume element. However, due to the fact that in this thesis, before the publication of the cited work, we had developed methods using the original formulation, we stick to the mathematical formalisms previously proposed by [Descoteaux et al. \[2007a\]](#).

Due to the need of a denser angular sampling, the ODF, $\Psi(\cdot)$, is estimated at the cost of longer acquisition times than DTI. The angular sampling is obtained through High Angular Resolution Diffusion Magnetic Resonance Imaging (HARDI) [\[Tuch et al.: 2002\]](#). This technique acquires a large number of DW images by performing a discrete sampling of points on the sphere as illustrated in ?? ??, and acquiring on each of the directions obtained. Hence, the HARDI signal at each voxel is a discrete spherical function, $S(\mathbf{g})$, with no a priori assumption about the nature of the diffusion process within the voxel. [\[Tuch: 2004\]](#) showed that an estimation of the ODF can be

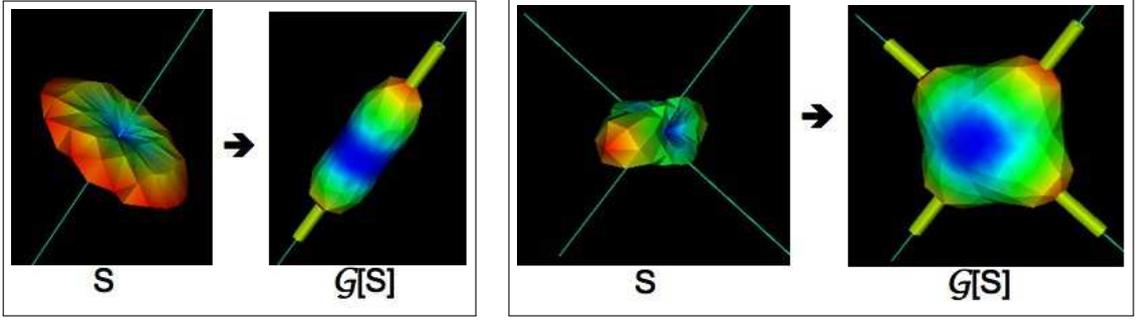


Figure 2.18: Funk-Radon Transform illustrated for the input diffusion attenuation signal S ($b = 1000$ s/mm^2) with 1 fibre (left) and two orthogonal fibres (right). The thin lines are the true underlying fibre directions and the thicker tubes are the detected maxima. One must imagine these functions as living on the surface of the sphere. Here, for visualization purposes, the radius of the respective spheres are scaled by the corresponding value on the surface. Blue to red colors represent low to high spherical values.

performed directly on the HARDI signal $S(\cdot)$ by applying the Funk-Radon transform $\mathcal{G}[\cdot]$:

$$\Psi(\mathbf{r}) = \mathcal{G}[S](\mathbf{r}) = \int_{S^2} \delta(\mathbf{r}^\top \mathbf{s}) S(\mathbf{s}) d\mathbf{s} \quad (2.11)$$

Intuitively, the Funk-Radon transform of a value at an arbitrarily defined “pole” of the signal on the sphere, is the integral of the spherical function S over the corresponding “equator” or great circle perpendicular to that “pole”. This is illustrated in figure 2.18.

Several analytical solution of the Funk-Radon transform have recently been proposed [Descoteaux et al.: 2007a, Anderson: 2005, Hess et al.: 2006]. These solutions are all based on the spherical harmonics representation of the HARDI signal, which gives a compact representation of the ODF.

As mentioned before, QBI [Tuch: 2004] reconstructs the diffusion ODF directly from the HARDI measurements on a single sphere by the Funk-Radon transform (FRT). The ODF is intuitive because it has its maxima aligned with the underlying population of fibres. However, computing statistics on a large number of discrete ODF values on the sphere² is computationally heavy and infeasible to integrate into a segmentation algorithm of the whole brain. A more compact representation of the ODF is thus needed. [Descoteaux et al.: 2007a, Anderson: 2005, Hess et al.: 2006] proposed a simple analytical spherical harmonic (SH) reconstruction of the ODF.

Letting Y_ℓ^m denote the SH of order ℓ and degree m ($m = -\ell, \dots, \ell$), we define a modified SH basis that is real and symmetric. For $\ell = 0, 2, 4, \dots, \ell_{\max}$ and $m = -\ell, \dots, 0, \dots, \ell$, we define a single index j in terms of ℓ and m such that

2. Typically, the ODF is reconstructed on more than 200 discrete spherical values.

$j(\ell, m) = (\ell^2 + \ell + 2)/2 + m$. The modified basis is given by

$$Y_j = \begin{cases} \sqrt{2} \cdot \operatorname{Re}(Y_\ell^m), & \text{if } m < 0 \\ Y_\ell^m, & \text{if } m = 0 \\ \sqrt{2} \cdot \operatorname{Im}(Y_\ell^m), & \text{if } m > 0. \end{cases} \quad (2.12)$$

where $\operatorname{Re}(Y_\ell^m)$ and $\operatorname{Im}(Y_\ell^m)$ represent the real and imaginary parts of Y_ℓ^m respectively. The basis is designed to be symmetric, real and orthonormal because of the normalization factor $\sqrt{2}$. Moreover, note that there are exactly $L = (1/2)(\ell_{\max} + 1)(\ell_{\max} + 2)$ terms in the spherical harmonic series of order ℓ_{\max} .

It is possible to obtain an analytical QBI solution [Descoteaux et al.: 2007a], final ODF Ψ , with

$$\Psi(\theta, \phi) = \sum_{j=1}^L \underbrace{2\pi P_{\ell(j)}(0) c_j}_{f_j} Y_j(\theta, \phi), \quad (2.13)$$

where θ, ϕ obey physics convention ($\theta \in [0, \pi], \phi \in [0, 2\pi]$), $L = (\ell + 1)(\ell + 2)/2$ is the number of elements in the spherical harmonic basis, c_j are the SH coefficients describing the input HARDI signal, $P_{\ell(j)}$ is the Legendre polynomial of order $\ell(j)$ ³ and f_j the coefficients describing the ODF Ψ . Here, we use the solution presented by Descoteaux et al. [2007a] with a Laplace-Beltrami regularization of the SH coefficients c_j to obtain a more robust ODF estimation. The detailed implementation of the Laplace-Beltrami regularization and HARDI signal estimation is presented in the work previously cited reference.

Distances between ODFs Once the ODF are computed, we want to capture similarities and dissimilarities between two ODFs, i.e two spherical functions $\Psi, \Psi' \in \mathbf{S}^2$ that can be represented by real SH vectors of length L , $f = \{f_1, \dots, f_L\}$ and $f' = \{f'_1, \dots, f'_L\} \in \mathbb{R}^L$, as shown in the previous section equation equation 2.13. Since the ODFs come from real physical diffusion measurements they are bounded and form an open subset of the space of real-valued \mathcal{L}^2 spherical functions with an inner product \langle, \rangle defined as

$$\begin{aligned} \langle \Psi, \Psi' \rangle &= \int_{\Omega} \Psi(\theta, \phi) \cdot \Psi(\theta, \phi)' d\Omega \\ &= \int_{\Omega} \left(\sum_{i=1}^L f_i Y_i(\theta, \phi) \sum_{j=1}^L f'_j Y_j(\theta, \phi) \right) d\Omega, \end{aligned} \quad (2.14)$$

where Ω denotes integration over the unit sphere. Because of the orthonormality of the spherical harmonic basis, $\int_{\Omega} Y_i(\theta, \phi) Y_j(\theta, \phi) d\Omega = \delta_{ij}$, the cross terms cancel and

3. $\ell(j)$ is the order associated with the j^{th} element of the SH basis, i.e. for $j = 1, 2, 3, 4, 5, 6, 7, \dots$ $\ell(j) = 0, 2, 2, 2, 2, 2, 4, \dots$

the expression is simply

$$\langle \Psi, \Psi' \rangle = \sum_{j=1}^L f_j \cdot f'_j. \quad (2.15)$$

Therefore, the induced \mathcal{L}^2 norm $\|\Psi\| = \sqrt{\langle \Psi, \Psi \rangle}$ giving us the distance metric between two ODFs is

$$\|\Psi - \Psi'\| = \sqrt{\int_{\Omega} (\Psi(\theta, \phi) - \Psi'(\theta, \phi))^2 d\Omega} = \sqrt{\sum_{j=1}^L (f_j - f'_j)^2}. \quad (2.16)$$

The Euclidean distance was also used successfully for DTI segmentation in [Lenglet et al.: 2006a] even though more appropriate metrics exist such as the J-Divergence [Wang et al.: 2004, Lenglet et al.: 2006b] and Riemannian geodesic distances [Lenglet et al.: 2006b]. Similarly, one can think of choosing another metric to compare ODFs. For instance, since the ODF can be viewed as a probability distribution function (PDF) of fibre orientations, one can use the Kullback-Leibler distance between two PDFs, as done in [Tuch: 2004]. However, in that case the problem quickly blows up computationally because one needs to use all N discrete HARDI data on the sphere instead of the L SH coefficients ($L \ll N$).

2.4.6 Beyond QBall-MRI

Techniques which model better the propagator like QBI and other methods based in HARDI made a great impact in the field of diffusion MRI. However, there are several shortcomings of these techniques that should be solved before its use in clinical settings becomes part of the routine. Firstly, HARDI acquisition schemes take long time, in setting where imaging time is critical, this is major difficulty. Moreover, there are no invariant measures of QBall imaging with a clear biological interpretation and an extensive use in applied scenarios validating their capabilities of serving as biomarkers for diseases or differences in brain capabilities.

The quest for new models, closer to the general case of the diffusion propagator while retaining the simplicity of use that characterizes the diffusion tensor, is a very active area of research. We refer the reader interested in such models to other work done in this direction by Jansons and Alexander [2003], Tuch [2004], Tournier et al. [2004], Wedeen et al. [2008], Aganj et al. [2009], Tristán-Vega et al. [2009], Ghosh and Deriche [2009] and Özarslan et al. [2009].

2.4.7 Summary

We have just described a brief although by no means exhaustive overview of the history of diffusion NMR particularly DTI and QBI. In this history, we have seen that DWI and in particular DTI are fit tools to characterize anisotropy in living tissue. Moreover, its usefulness in elucidating the microstructural features of fibrous

tissues, such as the human nervous system and more specifically, the brain's white matter, has been proven extensively. These microstructural features have also been used to explore neural function and neurological disorders [Jones: 2008, Catani: 2006, Beaulieu: 2002, Kubicki et al.: 2007]. However, up to this point we have only analysed what happens within a voxel in DWI and DTI. If we take a close look at figure 2.16b we may note that in areas where the tissue is highly organized, there seems to be a smooth continuity of the preferred diffusivity direction following the white matter structure. This observation is the kick-off idea to use DT-MRI as a tool to trace axon bundles in the white matter with a group of techniques called tractography.

2.5 TRACTOGRAPHY

Up to this point we have a representation of the water diffusion on each voxel of a DMRI image taken from the human brain. However, as we have seen in chapter 2, the pathways of the axons in the white matter are of the utmost importance in analysing the anatomy of the brain. Historically, these pathways were identified through dissection or traced using staining techniques directly on the brain tissues.

Diffusion MRI provides a unique, non-invasive set of tools to trace axonal pathways in the living brain. Either by following the preferred diffusivity direction or by using a probabilistic models of the diffusivity to simulate water particles moving in the brain, it can recover the connectivity and pathways of the brain's white matter. Diffusion MRI-based tractography can be divided in two families: global and streamline. Global, or probabilistic, tractography provides a connectivity measure between two regions of the brain. Even if it has been used to parcellate cortical [Anwander et al.: 2007] and sub-cortical [Behrens et al.: 2003] areas, the interpretation of these measures and their usability to find differences in neurological disorders still an issue [Jones: 2008]. On the other hand, streamline, or deterministic, tractography, although having its limitations [Johansen-Berg and Behrens: 2006], has been intensively studied and used in clinical and neuroscientific scenarios. In this section, we describe streamline tractography and briefly review its applications.

2.5.1 The streamline

In regions where the diffusion anisotropy is high, it is reasonable to assume that the major eigenvector of the diffusivity tensor is parallel to the local direction of the tracts [Beaulieu: 2002]. In these regions, a vector field can be obtained from the diffusivity tensor field, where the vector at each point is the major eigenvector of the tensor. Due to the previous arguments, any line with a tangent parallel to the vector field is a plausible estimation of a segment of the underlying fibre tract. Such lines are called streamlines.

Algorithm 2.1 Streamline tractography method with Euler integration [Basser et al.: 2000]

Input: A tensor field $\Sigma(\cdot)$ and seed point \mathbf{p}_0

Stopping criteria: Two thresholds: $t_{\text{anisotropy}}$ and $t_{\text{curvature}}$

Output: A trajectory $\mathbf{r}(\cdot)$ representing a white matter fibre bundle

```

1:  $s \leftarrow 0$ 
2:  $\mathbf{r}(s) \leftarrow \mathbf{p}_0$ 
3: repeat
4:    $\mathbf{r}(s + \Delta s) \leftarrow \mathbf{r}(s) + \Delta s \mathbf{e}_1(\Sigma(\mathbf{r}(s + \Delta s)))$ 
5:    $\text{anisotropy} \leftarrow \text{FA}(\Sigma(\mathbf{r}(s + \Delta s)))$ 
6:    $\text{curvature} \leftarrow \frac{\mathbf{e}_1(\Sigma(\mathbf{r}(s))) \cdot \mathbf{e}_1(\Sigma(\mathbf{r}(s + \Delta s)))}{\|\mathbf{e}_1(\Sigma(\mathbf{r}(s)))\| \|\mathbf{e}_1(\Sigma(\mathbf{r}(s + \Delta s)))\|}$ 
7:    $s \leftarrow s + \Delta s$ 
8: until  $\text{anisotropy} > t_{\text{anisotropy}}$  or  $\text{curvature} > t_{\text{curvature}}$ 
9: return  $\{\mathbf{r}(0), \mathbf{r}(\Delta s), \mathbf{r}(2\Delta s), \dots, \mathbf{r}(s)\}$ 

```

Mathematically speaking, in this case, a streamline is a 3D curve in space [Mori et al.: 1999, Basser et al.: 2000]. Letting $\mathbf{r}(s)$ be the point (x, y, z) , which is distance s along the streamline, the equation describing the evolution of the streamline is

$$\frac{d\mathbf{r}(s)}{ds} = \mathbf{e}_1(\Sigma(\mathbf{r}(s)))$$

where $\Sigma(\mathbf{r}(s))$ is the tensor at $\mathbf{r}(s)$ and $\mathbf{e}_1(\Sigma(\mathbf{r}(s)))$ is the unit length vector parallel to the eigenvector corresponding to the largest eigenvalue of $\Sigma(\mathbf{r}(s))$.

There are several algorithms and techniques available to solve the previous equation. In fact, it is a very active area of research. Therefore, we present only the basic equation and its major problems and then mention a set of representative publications which propose significant approaches to solve the problems. Due to the fact that none of these approaches has been extensively used in clinical applications, we will leave a more in-depth study of each algorithm to the interested reader.

2.5.2 Tracking the streamline

Revisiting the previous section, streamline tractography consists of tracking the trajectory $\mathbf{r}(s)$ of a white matter fibre tract from a field which comprises information about the water diffusivity on the brain. In the case of diffusion tensor streamline tractography, it consists of taking a tensor field $\Sigma : \mathbb{R}^3 \rightarrow PSD^3$ and an initial point in space, $\mathbf{p}_0 \in \mathbb{R}^3$, and clamping the start of the trajectory to that point $\mathbf{r}(0) = \mathbf{p}_0$. Then, the axonal trajectory is tracked through solving the differential equation [Conturo et al.: 1999, Mori et al.: 1999, Basser et al.: 2000]

$$\frac{d\mathbf{r}(s)}{ds} = \mathbf{e}_1(\Sigma(\mathbf{r}(s))). \quad (2.17)$$

Algorithm 2.1 solves equation 2.17 as proposed by Basser et al. [2000].

There are two main points to note about equation 2.17 and its proposed solution, algorithm 2.1. In the first place, it is a differential equation meaning that approximation error accumulates as the tracking process moves forward estimating the streamline. Secondly, the right hand side of the equation is defined on a *continuous diffusion tensor field*. Diffusion tensor MRI provides us with an imaging grid, a sampling of the field $\Sigma(\cdot)$, making interpolation on a tensor field necessary in order to solve equation 2.17.

Streamline tractography has had a remarkable success in performing *in vivo dissection* [Catani et al.: 2002] and obtaining biomarkers for neurological disorders [Ciccarelli et al.: 2008], however it is susceptible to errors. The causes of these errors can be characterized in three categories:

1. Noise in the diffusion image which causes a poor estimation of dominant diffusion directions.
2. Error in the modelling of the diffusion characteristics. This is a result of the underlying complexity of the white matter as opposed to the diffusion models which do not model complex configurations accurately.
3. Integration errors introduced in the tractography process by the algorithms chosen to solve equation 2.17.

A great number of new tractography algorithms are being proposed every year. Each algorithm aims to overcome some or all of these problems through a wide variety of mathematical tools. Here we briefly survey the most representative approaches and their characteristics:

- Conturo et al. [1999], Mori et al. [1999] and Basser et al. [2000] proposed variations of algorithm 2.1 which implement different approaches of numerical integration. This has an effect on the robustness of the algorithm to high curvature and dispersion [Lazar and Alexander: 2003].
- Westin et al. [2002] proposed a method based on tensor deflection in order to traverse regions of low anisotropy. However, this method proved to be less effective in regions where the tensor field had a definite principal direction than previous methods [Lazar and Alexander: 2003].
- Lenglet et al. [2004] and Fillard et al. [2007] addressed the problem of tensor interpolation by defining it on the Riemannian space of the symmetric positive 3×3 matrices. They achieved smoother tractographies and were capable of tracking fibres with higher curvature and through regions which were close to low diffusion anisotropy areas.
- Qazi et al. [2009] use a model where up to two tensors are admitted at each voxel. Hence, there are up to two possible principal directions that the streamline tracking can follow. This algorithm shows encouraging results in complex areas like the fanning of the CST to different areas of the motor cortex at the expense of a complicated interpolation of the diffusion field needed to solve equa-

tion 2.17. The question of how many tensors (principal directions) should be used at each voxel is still open.

- Descoteaux et al. [2009a] take advantage of HARDI models using QBall fields in order to perform a more accurate tractography in fanning regions allowing an arbitrary number of directions to be chosen. Even if there is a single seed, when the tracking reaches a fanning point, new seeds for streamline tracking are spanned in a dense sample of probable directions. This method produces a set of fibres which recovers segments of white matter fibres more accurately but does not reflect the continuity of the fibres or the cortical and subcortical regions they connected since they need to be cut in order to perform the splitting.
- [Malcolm et al.: 2010] fit local fibre orientations at each voxel while they perform the tractography using an unscented Kalman filter. This varies from all other techniques as they fit the model before performing the tractography. This approach is useful for obtaining smooth fibres in regions with complex diffusion patterns. Their work looks promising, however it lacks an in-depth validation study to fully grasp its value.

Despite the availability of streamline tracking algorithms, only a few of them, mostly the earliest and most studied ones [Conturo et al.: 1999, Mori et al.: 1999, Basser et al.: 2000, Fillard et al.: 2007], have made recurrent appearances in clinical research [Ciccarelli et al.: 2008]. This may be due to their efficacy in tracking major white matter structures, ease of implementation, availability in easy-to-use software packages or the existence of a study that clearly sets the characteristics of the algorithm for a clinician who must adjust the parameters [Lazar and Alexander: 2003].

In the following sections, we survey the main achievements of diffusion tensor streamline tractography (DTST) [Johansen-Berg and Behrens: 2009], the most important validation methodologies and their outcome, and its use in the search for biomarkers of neurological disorders.

2.5.3 In vivo dissection

Probably the first and most visually appealing result obtained from streamline tractography is its capacity to isolate and delineate major white matter fibre pathways or *in vivo dissection* of the white matter as in the dissection presented in figure 2.19. Stieltjes et al. [2001] and Catani et al. [2002] coined this term. Stieltjes et al. [2001] produced a dissection of the human brainstem. They performed an interrater reliability study in order to assess the reproducibility of their results and validated the dissection anatomically by performing qualitative comparison with histological data and atlases of the human nervous system. Catani et al. [2002] extended the previous work by carefully dissecting several white matter structures such as the Superior Longitudinal Fasciculus, the Internal Capsule and the Fornix among others. Catani et al. did not provide validation through qualitative or quantitative comparisons with other studies, however they performed an impressive job by

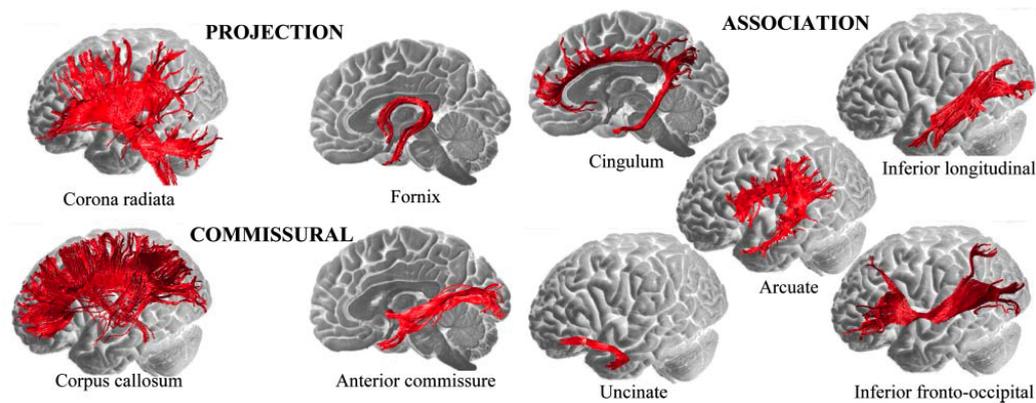


Figure 2.19: In vivo dissection using diffusion tensor streamline tractography. Major white matter tracts are shown dissected in accordance with anatomical knowledge. Adapted from [Catani et al. \[2002\]](#).

showing accurate reconstructions of each analysed structure and most of its components. Following the same line of work, [Wakana et al. \[2004\]](#) produced an atlas of the white matter based on tractography, a procedure which was later mimicked by [Catani and de Schotten \[2008\]](#). [Wakana et al.](#) showed the reconstructions of several white matter structures and provided a detailed description of how these structures were reconstructed establishing the basis for a reproducible method. Later, they provided a quantitative analysis of the reproducibility of *in vivo dissection* showing that it can be repeated with a high chance of success [[Wakana et al.: 2007](#)].

In vivo dissection of the white matter through streamline tractography produces appealing results which can be used in order to elucidate brain function and search for biomarkers of neurological disorders. However, it requires a considerable expertise in seeding regions on the brain to obtain white matter fibre bundles and then pruning these results to get rid of unwanted fibres in the bundle. This opens the quest for automatic *in vivo* dissection of the brain white matter.

2.5.4 Validation

Even if it has been shown that (diffusion tensor-based) tractography is a reproducible technique to reconstruct white matter structures from diffusion MRI, the question of its anatomical accuracy is still open, however. Several works have been done in order to address this problem in different settings.

Software Phantoms

The simplest approach to validation is the use of software phantoms. Software phantoms [[Alexander et al.: 2002](#), [Tuch: 2004](#), [Descoteaux et al.: 2007b](#)] were used frequently in order to validate different kinds of diffusion weighted sequences and diffusion models by simulating underlying tissues of different complexities. Par-

ticularly, these phantoms have been used in order to show limitations of the standard diffusion tensor model and prove that other models were more appropriate for these tissue configurations. However, most of these models are based on the diffusion properties of a single voxel which is not enough to validate tractography algorithms. With this in mind, [Lori et al. \[2002\]](#), [Lazar and Alexander \[2003\]](#), [Gössl et al. \[2002\]](#) and [Leemans et al. \[2005\]](#) proposed software phantoms based on global diffusion properties of white matter fibres instead of the tissue in one voxel. These phantoms were used in order to compare fibre tracking algorithms and evaluate their limitations. The result of these comparisons was that up to the year 2005, when HARDI models had not been extensively used for streamline tracking, the diffusion tensor-based streamline tracking (DTST) algorithms proposed by [Basser et al. \[2000\]](#) as the Runge-Kutta based method was the most effective and precise, except for the case where the fibre should bypass areas of low diffusion anisotropy where the tracking algorithm would be interrupted. Later, [Descoteaux et al. \[2009a\]](#) and [Malcolm et al. \[2010\]](#) used software phantoms in order to show how DTST algorithms react when bypassing areas of crossing and kissing fibres and to validate new tractography algorithms designed to overcome these limitations by using more complex models of the underlying diffusion pattern.

Software phantoms provide an inexpensive and highly controllable environment to validate tractography algorithms. However, the leap between these simulations and real data obtained through diffusion MRI is not negligible. It is hard to accurately simulate the noise and artifacts included by the imaging equipment and tissue characteristics such as permeability. Therefore, these phantoms provided an oversimplified yet useful set of scenarios in order to perform preliminary tests in a highly controlled setting with an available ground truth.

Physical Phantoms

In order to provide a controlled environment to validate tractography algorithms where different acquisition conditions are closer to real scenarios, physical phantoms were proposed. These phantoms can be divided into two families: synthetic and biological.

Synthetic Phantoms are built from synthetic fibres from medical applications as hemodialysis fibres [[Perrin et al.: 2005](#)] or textile fibres such as rayon, cotton [[Scifo et al.: 2004](#)], glass capillaires [[Yanasak and Allison: 2006](#)] or acrylic [[Poupon et al.: 2008](#)], provide a controlled environment for validating tractography algorithms. The synthetic fibres which form these phantoms are put together such that they generate different configuration and immersed in water and a relaxation agent [[Tofts et al.: 2000](#)]. A diffusion weighted MRI acquisition is then performed on these phantoms. The obtained images are used in order to validate diffusion models [[Poupon et al.: 2008](#)] and tractography algorithms [[Fillard et al.: 2009a](#)]. The main issues

in building these phantoms are the material, how to hold all the synthetic fibres together and which liquid medium should the phantom be immersed in. Summarizing several works edited by [Fillard et al. \[2009a\]](#), a key step in achieving a valid diffusion tensor based tractography is the previous regularization of the diffusion images [[Poupon et al.: 2000](#), [Fillard et al.: 2007](#), [Tristán-Vega and Aja-Fernández: 2009](#)]. These phantoms provide a valuable means to validate tractography algorithms and diffusion models. However, differences with microscopical geometry of brain tissue and its permeability characteristics limits the degree of validation provided by these methods.

Biological Phantoms can be made of muscle [[Basser et al.: 1994b](#)], excised and overlapped tissue from areas where the white matter has well-known configurations like the spinal cord [[Schwartz et al.: 2005](#), [Campbell et al.: 2005](#)] or the optic chiasm [[Roebroek et al.: 2008](#)]. Skeletal muscle has thicker fibres and simpler configurations than the brain's white matter. Due to these characteristics, it is suited to provide validation of streamline tractography algorithms in simple scenarios [[Damon et al.: 2002](#), [Heemskerk et al.: 2005](#)]. These studies show that tractography agrees not only with local orientations of the muscular tissue but with the global structure of the muscular fibres. [Descoteaux et al. \[2009a\]](#) used a phantom built of the combination of two excised and overlapped rat spinal cords [[Campbell et al.: 2005](#)] in order to validate a Q-ball based streamline tractography algorithm. [Descoteaux et al.](#) showed that this algorithm is capable of traversing crossing areas. Finally, [Roebroek et al. \[2008\]](#) performed a high spatial resolution acquisition at 9.4 teslas of four excised human optic chiasms. [Roebroek et al.](#) validated that, even at high resolutions, diffusion tensor-based tractography experiences problems in areas of complex configuration. This work cites more advanced methods like the ones considering a Riemannian geometry of the diffusion tensor manifold [[Lenglet et al.: 2004](#), [Fillard et al.: 2007](#)] and HARDI methods [[Descoteaux et al.: 2009a](#)] as candidates to overcome these problems.

Overall, biological phantom-based studies validate streamline tractography in several scenarios and show its limitations in complex tissue areas in detail. Although these studies provide a most valuable way to validate streamline tractography, the main limitation of these methods are the changes on the water diffusion properties of the tissue after it is excised and the simple architecture of the muscular tissue and the crossing produced by overlapping two rat spinal cords. Nevertheless, the use of these phantoms provides a valuable means to validate tractography in simple scenarios like muscular based cases, and in scenarios that could not be reproducible in clinical environments due to high magnetic field or long acquisition times.

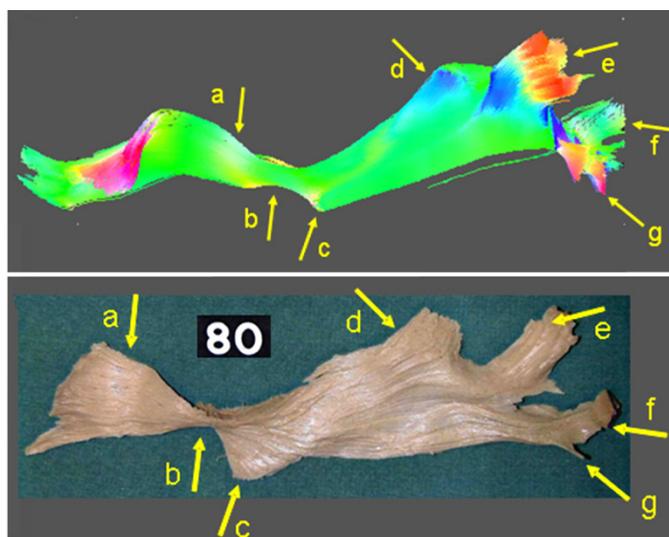


Figure 2.20: Validation of *in vivo* dissection by comparing it to gross dissection. Adapted from Lawes et al. [2008]. The left inferior fronto-occipital tract (IFO) is shown. Top, IFO obtained through *in vivo* dissection. Bottom, IFO obtained from gross dissection. The letters indicate areas of similarity. For a detailed description of these areas see Lawes et al. [2008].

Anatomical comparison

Work has also been done in order to validate streamline tractography against human and other mammal brains. Usual validation is done through tracking a well known white matter fibre bundle like the corpus callosum, the cortical spinal tract [Basser et al.: 2000] or the optic chiasm [Roebroek et al.: 2008] much like the *in vivo* dissection procedure mentioned in subsection 2.5.3. However, these studies do not provide a comparison with the actual subject being studied. Lawes et al. [2008] overcame this by performing a remarkable study in order to qualitatively match the results of streamline tractography with gross dissection, see figure 2.20. With the aim of providing quantitative validation, Dauguet et al. [2007] and Schmammann et al. [2007] acquired diffusion weighted images of a macaque brain and then used chemical tracers to compare *in vivo* tractography results with stained histological sections of the post-mortem brain. They showed general agreement between both of these techniques. However they also stressed the dependence of the results with the fine-tuning of the of the tractography algorithm's parameters and the problems derived from the registration of the *in vivo* and post-mortem modalities.

2.5.5 Limitations

Up to this point, we have reviewed streamline tractography methods and several of the studies that have been used for its validation. These studies show that tractography constitutes a useful tool in tracing white matter pathways in most cases, however it has a number of limitations [Johansen-Berg and Behrens: 2006].

In the first place the size of the imaging voxels recovering the diffusion information is typically a few cubic millimetres, hence a voxel can contain tens of thousands of axons [Beaulieu: 2002]. The majority of the tractography methods assume that the underlying tract is well described by a single orientation estimate. There is an active area of research aiming to solve this through developments in spatial [Jaermann et al.: 2004, Peled and Yeshurun: 2001] and orientational [Tuch et al.: 2002] resolution, diffusion modelling [Descoteaux et al.: 2007a, Tuch: 2004, Tournier et al.: 2004, Aganj et al.: 2009, Tristán-Vega et al.: 2009] and tractography [Malcolm et al.: 2010, Hageman et al.: 2006, Deriche and Descoteaux: 2007, Jbabdi et al.: 2004, 2007, Iturria-Medina et al.: 2007]. Some of these: state-of-the-art techniques already produced sensible anatomical results [Tuch et al.: 2005, Cohen-Adad et al.: 2008, Perrin et al.: 2008]. Other fundamental limitations include the inability to differentiate afferent from efferent pathways, locate synapses or determine whether the fibre is active at a given time. Nevertheless, recent work seeks to overcome the latter limitations and obtain activation—hence functional—information out of diffusion weighted images [Le Bihan et al.: 2006, Aso et al.: 2009a].

To conclude, the limitations mean that tracts that are not present in the imaged tissue might appear after tractography (false positives) and that tracts that are present in the same tissue might not appear (false negatives). For this reason, results involving diffusion weighted-based streamline tractography should always be interpreted with care. Several studies have proven and validated that streamline tractography is capable of providing biomarkers of various neurological disorders. Tractography achieves this by supplying the means to quantitatively assess the location or presence of white matter tracts, defining paths over which quantitative comparison can be performed among populations or time and providing new information on normal anatomy relevant to surgical planning or an understanding of diseases.

2.5.6 Streamline tractography in neurological disorders

We have seen in this section that diffusion-based tractography is able to generate virtual, three-dimensional representations of the white matter fibre tracts. Moreover, we have seen that several studies concluded that, up to a certain degree of precision, tractography reconstructions are anatomically correct. This ability to trace the white matter pathways in a non-invasive way has generated enthusiasm and high expectations in the medical community [Mesulam: 2005].

In recent years, the amount of medical-oriented applications of tractography underwent an explosion. In this section, we survey some of the most relevant of these works following the reviews of Ciccarelli et al. [2008] and Kubicki et al. [2007]. From these reviews we selected only the studies that introduce or clearly illustrate different measures obtained from diffusion tensor-based streamline Tractography (DTST) as biomarkers. For more extensive reading on disorders and, and the clinical interpretations of these biomarkers we refer the reader to the previously mentioned reviews.

Table 2.1 shows a summary of the tractography-based biomarkers presented in this section and the non-diffusion-based indices used to validate these biomarkers.

Stroke

Stroke is the first clinical application where the diffusion weighted imaging was used in order to visualize and quantify damage [Pierpaoli et al.: 2001]. Although its routine clinical application does not require tractography, several studies have used this technique in order to obtain quantitative measures along the pathways affected by the stroke.

Several different studies analysed the scalar quantities derived from DT-MRI along white matter bundles obtained through streamline tractography [Møller et al.: 2007, Konishi et al.: 2005]. These studies were able to relate the volume of the lesion with tract involvement [Konishi et al.: 2005] and correlate changes on scalar quantities with specific neurological scores [Møller et al.: 2007]. Changes of scalar quantities along the tracts are the most usual biomarkers analysed using diffusion tensor tractography.

Analysing less typical biomarkers, Cho et al. [2007] performed a study on 55 patients with one hand completely paralysed after a corona radiata stroke. They traced the left and right CST using DTST and divided the results into four types (figure 2.21):

Type A, the CST was preserved around the infarct, the tract originated in the affected hemisphere including the primary motor cortex and passing around the infarct to the medulla.

Type B, the CST was similar to type A, except the fibre originated in another cortex (primary sensory cortex, posterior parietal cortex, premotor cortex and pre-frontal cortex), but not the primary motor cortex.

Type C, the CST was interrupted at the infarct.

Type D, the CST did not reach the infarct due to degeneration.

Six months after the stroke, the motor function of the affected hand was evaluated with several indices of recovery for the affected hand measuring strength and finger movement. These showed significant correlation with the DTST type. Highest recovery scores were seen in the DTST type A patients; patients with lower motricity index had DTST type D. Cho et al. concluded that the integrity of the corticospinal tract determined by DTST obtained during the early stage of a corona radiata infarct was helpful in predicting the motor outcome of the affected hand. Hence, proving that the existence of a tract, if carefully analysed, is a valid biomarker.

Schaechter et al. [2008] analysed 10 patients after a stroke which resulted in loss of hand strength in one hand and 10 healthy controls. Their study showed that a decrease in the number of white matter fibre tracts of the CST obtained through

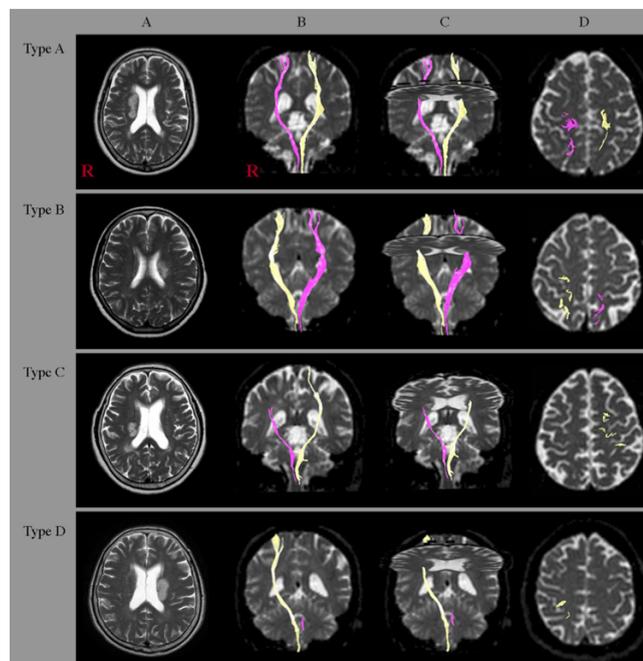


Figure 2.21: Diffusion tensor tractography (DTST) result types in stroke patients with impaired hand movement reproduced from [Cho et al. \[2007\]](#): Type A, normal CST after the infarct; Type B, similar to A, except that the affected CST reaches a different area of the cortex; Type C, CST interrupted at the infarct; Type D, CST did not reach the infarct due to degeneration.

tractography on the hemisphere affected with the stroke correlates positively with increased activation on the ipsilesional motor area (M1/S1) (figure 2.22). In their work they argue that this demonstrates a quantitative relationship between the damage of the CST, measured as the quantity of fibres that could be traced, and cortical activation in stroke patients.

Multiple Sclerosis

Streamline tractography has been used in order to address four questions in Multiple Sclerosis:

- Can tractography be used to assess the pathological process in the white matter pathways?
- Does tractography improve the correlation between radiological findings and clinical disability?
- Can a tractography-derived connectivity index correlate with disability better than fractional anisotropy?
- Does tractography detect the structural changes that contribute to functional adaptive changes?

The first issue a MS tractography study has to deal with is the effect of the white matter lesions on tractography results. Specifically, there is a reduction of the fractional anisotropy index in the voxels which encompass the MS lesions. For this rea-

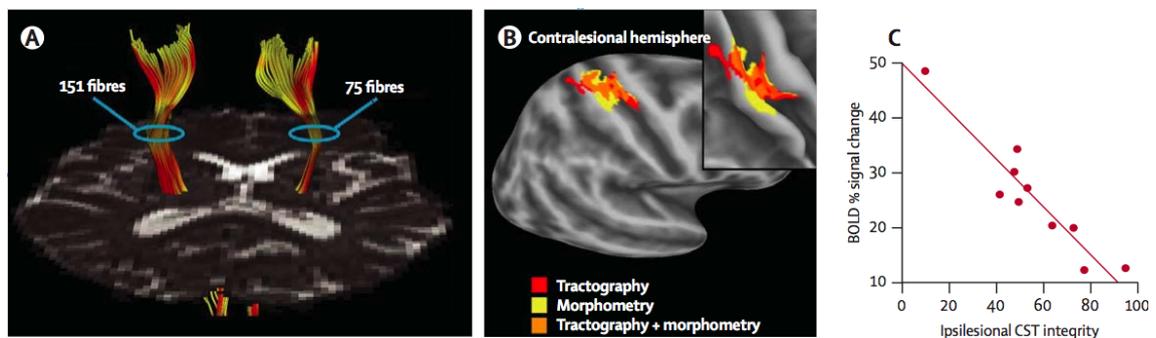


Figure 2.22: Corticospinal tracts reconstructed by diffusion tensor streamline tractography (A). Areas where fMRI shows morphological overlap with tractography reaching the M1/S1 cortical regions (B). Correlation between fMRI BOLD signal on the M1/S1 cortical regions and tractography integrity. Adapted from [Schaechter et al. \[2008\]](#) by [Ciccarelli et al. \[2008\]](#).

son, streamline tracking algorithms might stop the tracking procedure even when there is a white matter fibre traversing the lesion [[Lin et al.: 2007](#), [Pagani et al.: 2005](#)].

[Lin et al. \[2007\]](#) and [Pagani et al. \[2005\]](#) analysed around 30 MS patients with affected mean pyramidal functional system but normal-appearing white matter in non diffusion-weighted MR images and around 20 healthy control subjects. In order to overcome the problem of interrupted fibres while segmenting the pyramidal tract, they generated a template probability map of the pyramidal tracts in healthy subjects, as shown in figure 2.23a. Then, they analysed diffusivity indices on those maps. Both research groups analysed fractional anisotropy (FA), axial (λ_1), radial (λ_{\perp}) and mean (MD) diffusivity. They performed statistical analyses by comparing the distribution of these indices within the tract probability maps using the Mann Whitney-U test for non-Gaussian distributions and correlated these indices with the volume of the MS lesion. Both groups had similar findings of significant differences in the four analysed indices and argued the clinical significance of those indices in the diagnosis of MS, relating these differences with a Wallerian degeneration [[Pierpaoli et al.: 2001](#)] of the pyramidal tracts. Following their previous work, [Lin et al. \[2007\]](#) performed a more detailed study on patients with a different kind of MS, neuromyelitis optica. In addition to of the previous analysis, they normalized the length of the pyramidal tracts of the subjects and analysed the variation of the previously selected diffusivity indices along the tracts, as seen in figure 2.23b. They found that in this case the λ_1 had no significant variation between patients and controls. Moreover, they showed that while FA and λ_{\perp} were different all along the tract with the exception of the segments closer to the cortex, the MD showed significant differences along the length of the tract. They correlated this with previous work on the biological basis of diffusion and argued that there was an increased permeability in the axonal membrane or myelin sheath of the axons.

Instead of producing tract probability maps from streamline tracking, [Ciccarelli et al. \[2005\]](#) used a global tractography method by [Parker et al. \[2002\]](#) in order to tract through MS lesions. With a sample of 7 patients and 10 controls, they found that there were no changes in the fractional anisotropy while the voxels connected with the optical tract showed a small difference in connectivity with statistical significance. Although they admitted that their samples size was too small to be representative, they argued that the global tractography method from [Parker et al. \[2002\]](#) is to tract through lesions and detect differences in connectivity.

[Audoin et al. \[2007\]](#) combined tractography and fMRI in order to study structural plasticity on MS. Their sample consisted of 24 early MS patients with affected working memory and 15 healthy controls. After using a public atlas to select a set of cortical and subcortical regions involved in working memory, they traced the white matter fibres joining them and constituting the executive system of working memory. Regions that were untraceable due to MS lesions were not considered. Then, using the Mann Whitney-U test to compare non-Gaussian distributions, they searched for differences in the distribution of MD and FA on the selected tracts. This analysis showed a significant increase of the MD on tracts connecting Brodmann's 45/46 and 9 areas on the right hemisphere and the left and right Brodmann 9 areas. Also, increased MD and decreased FA were found connecting Brocca's area and the thalamus on the right side. They measured connectivity by counting the fibres joining each pair of selected regions. Using the Wilcoxon rank test, a significant increase of connectivity between the thalami was found. Finally, their study showed correlations between the connectivity and diffusion indices and an fMRI study involving the selected cortical and subcortical regions. As a result of their study they claimed that the combination of DTST and fMRI detected structural changes on the working memory circuit and evidenced white matter plasticity.

2.6 SUMMARY

Diffusion tensor based streamline tracking provides the means to recover anatomical structures of the human cerebral white matter. An extensive set of studies have been performed in order to validate this technique and have exhibited several of its limitations. To overcome these limitations, a constantly growing set of new techniques exists, from the algorithms used to perform the tractography to the acquisition sequences. Nevertheless, most of these techniques are not easy to implement in clinical scenarios due to long acquisition times or the fact that the complex algorithms have a set of parameters which are not easy for non-expert users to tune.

Even if several limitations have been reported, several studies have also shown that this technique is reproducible and recovers the major white matter structures accurately. Moreover, this technique has been widely adopted by clinicians and neuroscientists. There is a growing body of clinical studies showing that diffusion tensor-

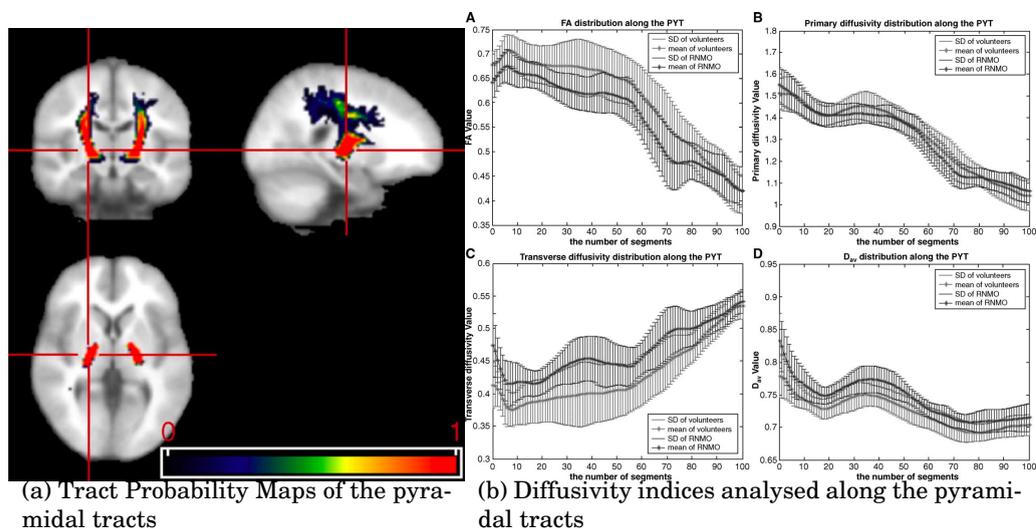


Figure 2.23: On the left (a): Tract probability maps for the pyramidal tracts generated from a set of healthy subjects over a T1 template. The colour scale indicates the probability that a fibre from the pyramidal tract traverses the region. These maps were generated in order to overcome limitations on tractography due to Multiple Sclerosis lesions. Adapted from Pagani et al. [2005]. On the right (b): diffusivity indices analysed along the tracts for MS patients compared to healthy controls. Adapted from Lin et al. [2007]

Table 2.1: Tractography-based biomarkers proposed in investigating brain stroke and multiple sclerosis. The table show the author of the work, the proposed biomarker and the non-diffusion-based index used to correlate and validate the tractography-based biomarker.

Stroke		
Author	Biomarker	Correlation Index
Konishi et al. [2005]	Statistical analysis of diffusivity indices along the tract	Volume of the lesion and tract involvement
Møller et al. [2007]	Statistical analysis of diffusivity indices along the tracts	Neurological scores
Cho et al. [2007]	Tract existence and reach	Paralysis recovery scores
Schaechter et al. [2008]	Number of fibres	Activation surface on the ipsilesional motor area
Multiple Sclerosis		
Author	Biomarker	Correlation Index
Ciccarelli et al. [2005]	Connectivity measure through global tractography	Previous anatomical knowledge of the disease
Pagani et al. [2005]	Averaged diffusivity indices on the tract probability map	MS symptom scores
Lin et al. [2006]	Statistical analysis of diffusivity indices along the tract	Comparisons with previous studies
Lin et al. [2007]	Averaged diffusivity indices on the tract probability map	MS lesion size
Audoin et al. [2007]	Averaged diffusivity indices on the tract probability map and number of tracts	Functional (fMRI) connectivity

based tractography is capable of providing biomarkers. These biomarkers have been shown to have a wide range of applications, from elucidating characteristics of the healthy brain to identifying characteristics of neurological disorders, contributing to its diagnosis. However, most of the clinical studies performed require heavy user interaction to identify major white matter structures and to analyse their characteristics. This calls for a tool which allows for the sound performance of statistics on white matter fibre tracts obtained through diffusion tensor-based tractography techniques and to automatically identify white matter structures.

CLUSTERING

Comparisons are odorous.

Much Ado About Nothing,
William Shakespeare, 1600

The main goal of this thesis is the automatic identification of white matter structures in the brain. Hence, given a set of elements which could be the voxels of a dMRI image or white matter fibres obtained through tractography, our goal is to group them such that these groups coincide with anatomical structures. Having said all this, we direct our attention to a particular research area in computer science which deals with the task of grouping elements, clustering.

3.1 WHAT IS CLUSTERING?

Data clustering was conceived in order to discover the natural grouping of a set of elements. The need to automatically group animals, diseases, rocks and almost any other examples of element or observation of has led to the development of clustering methodologies in most scientific areas dealing with recollection and processing of real data. The first article with the term “data clustering” in its title appeared in 1954 and dealt with anthropological data. Since then, there has been an enormous number of contributions to this discipline, averaging a minimum of least 1,300 articles dealing with data clustering per year during the last 10 years. This copious amount of work makes the task of reviewing the literature, colossal. Several surveys on data clustering have been published, the most relevant being the articles by Jain [Jain: 2009, Jain et al.: 1999, 2000] and the books written by Jain and Dubes [1988], Duda et al. [2001] and Hastie et al. [2009].

Despite this amount of work, clustering algorithms can be roughly divided into two main classes: *hierarchical* and *partitional*. Hierarchical algorithms find nested clusters in an agglomerative fashion or divisive fashion. Partitional algorithms on the other hand, find a partition of the data, looking for all the clusters simultaneously, and do not impose a hierarchical structure. The input of a hierarchical algorithm is an $N \times N$ matrix where N is the number of elements to be clustered. Each entry of this matrix quantifies the proximity between a pair of elements. A partitional al-

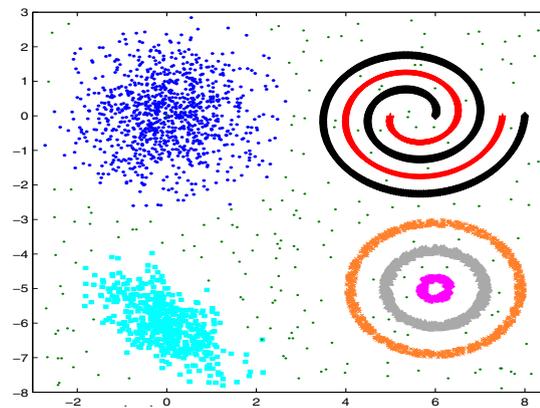


Figure 3.1: Seven clusters coloured as they should be grouped by an ideal clustering algorithm. None of the available clustering algorithms can detect all of these [Jain: 2009]. Image adapted from [Jain and Law: 2005]

gorithm, on the other hand, can either use a similarity matrix or an $N \times D$ pattern matrix, where N elements are embedded in a D -dimensional feature space. Using what is called a (dis)similarity function, the proximity matrix can be easily derived from the feature matrix. Moreover, there exists a series of algorithms which perform the converse task, usually called manifold learning techniques [Pearson: 1901, Torgerson: 1952, Scholkopf et al.: 1999].

Data clustering is typically used for the following purposes [Jain: 2009]:

Underlying structure: to gain insight into data, generate hypotheses, detect anomalies, and identify salient features.

Natural classification: to identify the degree of similarity among forms or organisms.

Compression: as a method for organizing the data and summarizing it through cluster prototypes.

Clustering, however, has a main problem: the definition of a cluster is somewhat fuzzy and very dependant on the application. In figure 3.1 different types of clusters are illustrated. The distinctive characteristics of these clusters are density, isotropy (blue), anisotropy (cyan), different scales of the same cluster (orange, grey, pink), non convexity (black, red), lack of linear separability and inclusion of clusters (orange, grey, pink and black, red). These are most of the issues clustering algorithms deal with, successfully or not.

An operational definition of clustering can be stated as follows: Given a representation of N objects, find K groups based on a measure of proximity such that the objects in the same group are close while objects in different groups are distant. However, we must address several questions in order to achieve this task:

Data representation: How do we define proximity between two objects?

- What features of the objects should we use to distinguish them?
- Should we normalize the proximity measure?

Cluster hypotheses: What are the expected characteristics of the clusters?

- Do we know the number of clusters we are expecting to find?
- Can we expect the clusters to be convex? Do we have cues as to their shapes?
- Is density a distinctive factor?
- Do the clusters have different scales?
- Is it possible that there are outliers?

Cluster validity: After the clustering process has been performed, are the clusters valid? Did the algorithm produce an output which is coherent with what we know of the data?

Studying the data and the analysis that we are expecting to perform using clustering is decisive in answering the previous questions.

3.1.1 Data representation

One of the most important factors influencing the performance of the clustering algorithm is data representation. This is the choice of features which represent each element to be clustered. If this representation is good, it is likely that the clusters that we are looking for are compact and isolated. Unfortunately, there is no universally good representation; it must be chosen using domain knowledge [Jain: 2009].

A simple example of how data representation simplifies a problem is shown in figure 3.2: two representations of the same set of points are shown in this image, rectangular and polar coordinate systems. Looking at the scatter plots of both representations drawn in Cartesian systems, the difference is remarkable. Analysing this case in the rectangular representation is difficult for most clustering algorithms because one cluster is contained in the other and due to their lack of convexity. On the other hand, the polar case presents two compact and isolated clusters which can be separated linearly, a trivial case. Moreover, there is an issue of scale in these two clusters: the larger circle has exactly the same shape as the smaller one, they are simply differentiated by their scale. When using the polar representation, this matter of scale disappears as the spread around the mean radius for both clusters is the same. Thus, we can affirm that the polar representation has a *normalizing* effect on this dataset, removing scale differences.

In cases where the analysed elements are more complex than points in space, the features chosen to represent each element are of great importance. As an example we use the classification and organization of animals presented by Pampalk et al. [2003]. As seen in figure 3.3a, we have a set of animals and two kinds of features, appearance and activities. It is possible to use each row as a Boolean vector usually called a feature vector. Then, we feed these feature vectors to a clustering or visualizing algorithm in order to group animals. In figure 3.3b we show a spatial organization of the animals based on their activities, the lighter the background colour, the higher the density of animals in that area. Conversely, we show a spatial or-

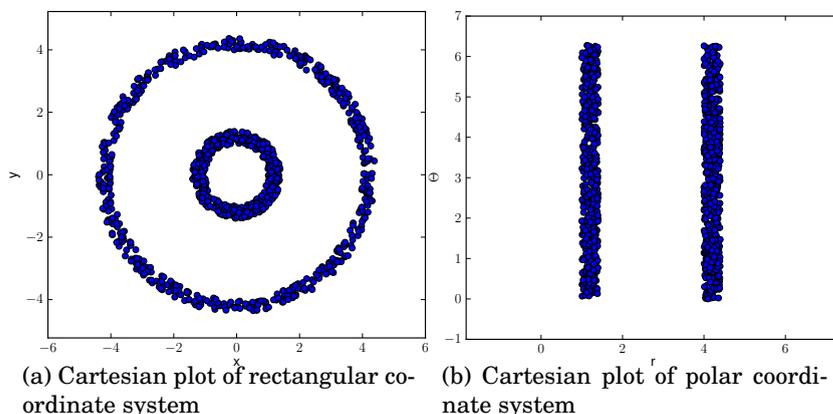


Figure 3.2: Two representations of the same set of elements to cluster in rectangular (a) and polar (b) coordinate systems. While the rectangular representation presents a difficult setting for most clustering algorithms, the polar representation is a trivial setting for all of them.

ganization based on their appearance in figure 3.3c. In figure 3.3b it is possible to observe that birds are pushed towards the lower part of the diagram, while mammals towards the upper part. An exception of this is the cow which seems to be either a strange type of bird or to be classified with the chicken, somehow as a farm animal. Moreover, predators have been pushed towards the left side while herbivores towards the right. In figure 3.3c where animals are arranged according to their activities, we can see a clear separation between birds and mammals. Within this arrangement animals go from smaller on the left, to bigger on the right. It is clear from the comparison of these two graphs that the choice of features leads to completely different results.

The representation of the data is of utmost importance to clustering algorithms. Different representations can not only turn a hard problem into an easy one (and the other way around) but they also play a role in the interpretation of the data which will be done after the clustering.

3.1.2 Representation Matrices

Clustering algorithms mostly deal with two types of data representation: a *pattern matrix* or a *proximity matrix*. An example of pattern matrix is shown in figure 3.3a. In this matrix there is a row for each element and a column for each feature. On the other hand, the proximity matrix is a squared matrix with as many rows (and columns) as elements in our dataset. If we notate this matrix as A , then the entry A_{ij} contains a number quantifying how close the elements i and j are, a proximity index. There are two kinds of proximity indices:

Similarity where A_{ij} is maximal when i and j are considered as indistinguishable and minimal when they are as different as possible. In this matrix $A_{ii} \geq A_{ij}$. A simple example of similarity measure between two vectors is the inner prod-

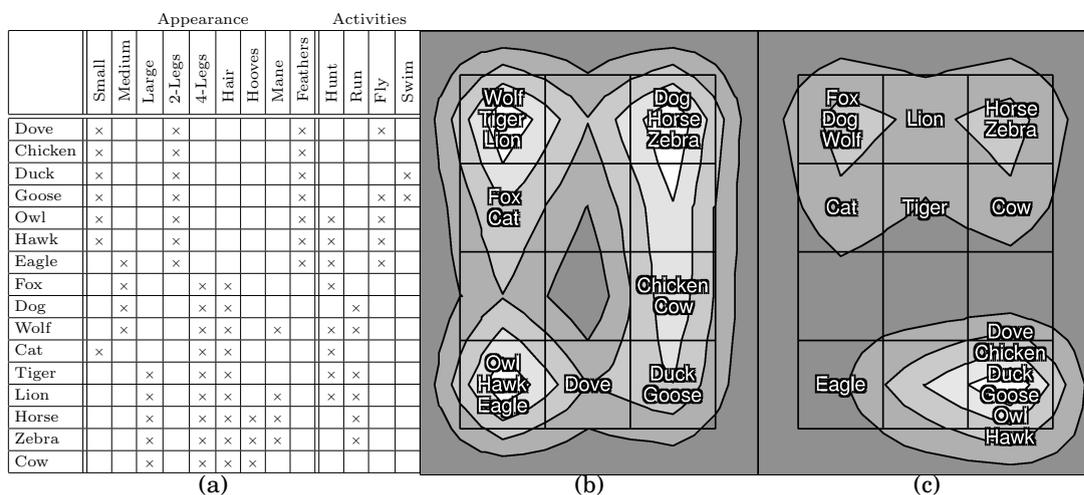


Figure 3.3: Animals spatially organized in relation to two types of features: (a); according to their activities (b); and based on their appearance in (c). The lighter the background the higher, the density when it is lighter. Adapted from Pampalk et al. [2003].

uct. Let \mathbf{v}_i and \mathbf{v}_j be two vectors, then, their inner product

$$A_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$$

is 0 when they are perpendicular and equal to $\|\mathbf{v}_i\|\|\mathbf{v}_j\|$ when they are parallel.

Dissimilarity where A_{ij} is minimal when elements i and j are considered indistinguishable and maximal when they are as different as possible. In this matrix $A_{ii} \leq A_{ij}$. Dissimilarity can be easily shown with the Euclidean distance between two vectors,

$$A_{ij} = \|\mathbf{v}_i - \mathbf{v}_j\|.$$

In this representation, $A_{ij} = 0$ if the two elements are the same and it can reach $A_{ij} = \infty$ if they are not similar at all.

It is important to note that the elements to be clustered do not need to be in an inner-product space (similarity case) nor in a metric space (dissimilarity case). Several of the proximity measures do not usually meet the hypotheses needed to constitute these spaces. However, if one is able to show that the elements “live” in a metric or inner-product space, inherent properties of these spaces are helpful in finding algorithmic optimizations [Jain and Dubes: 1988].

Knowing that data should carefully be analysed and represented in order to be clustered, we move to the analysis of the clusters that we are looking for.

3.1.3 Cluster Hypotheses

The first hypothesis that we can make concerns the structure of the clusters: are we looking for a hierarchy or for a partition? A hierarchy seems logical in cases like a clustering of human body parts: when the scenario is such that elements are

“little finger”, “right forearm”, “torso”, “right leg”, “big toe”, etc. and the fact that the parts are connected or not is the similarity measure. In cases like having to cluster food items as “carbohydrates”, “dairy products”, “vegetables”, “fruits”, etc. the goal is clearly to look for a partition and not a hierarchy.

In partitional clustering, the number of clusters is usually the next point to take into account. Generally, there is no way to know this number in advance. Even if the most popular clustering algorithm, K-Means [Jain: 2009], and the majority of partitional algorithms require this as a parameter, it is usually impossible to directly infer this information. Several approaches have been taken to automatically infer the number of clusters: statistical descriptors [Figueiredo and Jain: 2002, Hansen and Yu: 2001]; information theory-based approaches [Fraley and Raftery: 1998, Celeux and Soromenho: 1996]; and a wide range of mathematical tools such as eigenvalue analysis [Tibshirani et al.: 2001] or the formulation of an infinite mixture model [Rasmussen: 2000]. However, no method works in the general case.

The second hypothesis we analyse is in reference to the form. If we picture each element as a point in space, then the clusters should be given by high-density groups. As we can see in figure 3.1, given this graphical representation, the clusters may have different shapes. The first point to make regarding shape is whether the clusters are convex or not. Provided that the clusters are convex, the job of clustering algorithm is simpler. This is due to the fact that searching for a convex hull containing each cluster separately is a relatively easy task and the most usual clustering algorithms perform it successfully. If convexity can not be assured, the user of the clustering algorithm must analyse whether or not there is a family of cluster shapes. If there is, these shapes can be represented as parametric probability distributions and the clustering algorithm can be cast into an algorithm to find mixtures of probability distributions, like the Expectation-Maximization algorithm [Dempster et al.: 1977]. If it is not possible to make an assumption about the cluster shape, the user must resort to methods without cluster bias like mode seeking methods or specifically modified versions of other clustering techniques.

The last issue to consider is the variability of density or scale. Clusters might have densities which largely differ or, equivalently, have very different scales. In this cases, the algorithm must take that into account and most-likely normalize the proximity values according to density estimates. The occurrence of outliers is another characteristic consider. If outliers are expected, the algorithms must be robust to them in order to find valid clusters.

In the same way that data representation and proximity measures lead to different results, the choice of the clustering algorithm leads to different clusters for the same database. Therefore, it is fundamental to know the biases of each algorithm and how the obtained results are influenced by them.

3.2 CLUSTERING ALGORITHMS

Up to this point, we have reviewed the conditions and hypotheses which we should analyse before selecting or designing a clustering algorithm. In this section we present the most widely use approaches to clustering algorithms divided in categories. These categories are not definitive since several algorithms can be cast into more than one of them but they are useful in understanding the different approaches and their consequences. Taking all of this into account, we survey the most representative clustering algorithms of each family.

3.2.1 Hierarchical Algorithms

Hierarchical algorithms produce a nested series of partitions. This family is subdivided in two: *agglomerative* and *divisive*. Although divisive algorithms exist in theory, there are almost no examples of this algorithmic family in order to recover a hierarchy of the analysed set of elements [Jain: 2009, Jain and Dubes: 1988, Hastie et al.: 2009], hence we will not describe them in this section.

Agglomerative algorithms start from a scenario where every element is in its own cluster and join, recursively, the two most similar clusters until no other pair of clusters can be joined. The result of these algorithms is usually a tree or a set of trees (forest) in the case that of isolated sets of elements. The general scheme for this family is shown in algorithm 3.1. The algorithms of this family are usually variations of three techniques. Single, complete and average-linkage. The difference among these techniques is in how the dissimilarity between elements, $d(\cdot, \cdot)$, is generalized to a dissimilarity between clusters, $\mathbf{d}(\cdot, \cdot)$:

Single-linkage: inter-cluster similarity represents the distance between two clusters as the distance of their closest elements

$$\mathbf{d}(P, Q) = \min_{p \in P, q \in Q} d(p, q)$$

Complete-linkage: inter-cluster similarity uses the distance between the most distant elements of the cluster

$$\mathbf{d}(P, Q) = \max_{p \in P, q \in Q} d(p, q)$$

Average-linkage: inter-cluster similarity uses the mean distance between the two clusters

$$\mathbf{d}(P, Q) = \frac{1}{|P||Q|} \sum_{p \in P} \sum_{q \in Q} d(p, q)$$

The single-linkage algorithm has a tendency to produce elongated clusters while the complete and average-linkage ones tend to produce round, isotropic clusters. All of these algorithms are biased towards isotropic clusters. At the expense of adding parameters, modifications to these algorithms can make them apt to detect arbitrar-

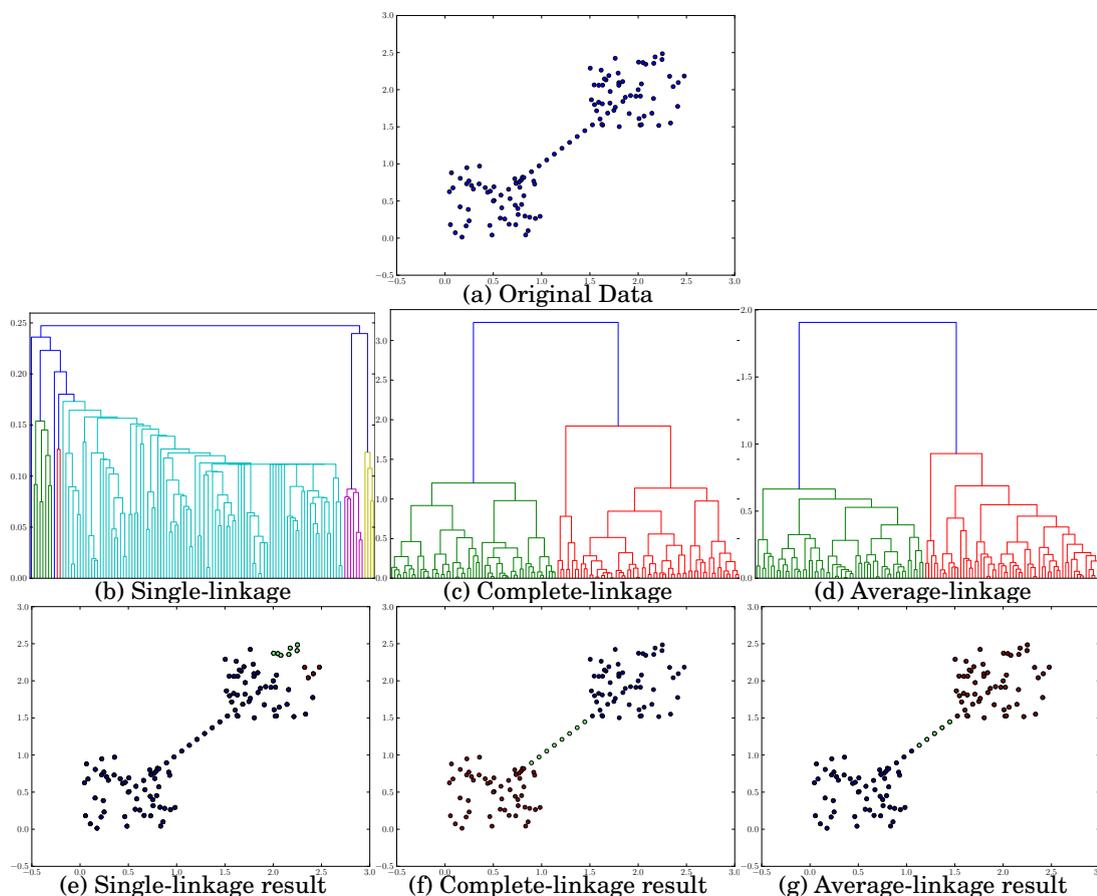


Figure 3.4: Agglomerative clustering results. Top row: original data. Middle row: dendrograms for a set of points clustered using the Euclidean distance as a dissimilarity measure. Bottom row: clusters obtained by thresholding the dendrograms to obtain 3 clusters.

ily shaped clusters, this is the case of the CHAMELEON hierarchical clustering algorithm proposed by [Karypis et al. \[1999\]](#).

An important property of agglomerative algorithms is *monotonicity*. That is, the dissimilarity (similarity) between merged clusters is monotone increasing (decreasing) with the level of the merger. Hence, the results of agglomerative clustering can be plotted as a tree such that the height of each node is proportional to the value of the intergroup dissimilarity $d(\cdot, \cdot)$. This kind of plotting is called a *dendrogram*, shown in figure 3.4. The dendrogram is a useful tool to visualize the hierarchy and to perform interactive exploration of obtained clusters. Moreover it is responsible for the success of the agglomerative algorithms [[Hastie et al.: 2009](#)].

Finally, if the user wants a set of clusters as a result instead of a hierarchy, several criteria can be applied to obtain those. The simplest technique is to find the highest nodes below (above) a threshold inter-cluster dissimilarity (similarity). This is shown in figure 3.4. Applying a threshold of 0.23 to the dendrogram in figure 3.4 leads to 3 clusters; A threshold of 1.5 leads to 2 clusters in the dendrogram shown in figure 3.4c and in the dendrogram of figure 3.4d, 0.8 leads to 3 clusters.

The monotonicity, ease of visualization and lack of parameters are two of the ma-

Algorithm 3.1 General algorithm for hierarchical agglomerative clustering

Inputs: A dissimilarity function between clusters, $d(\cdot, \cdot)$, and a set of elements X .

Outputs: A graph representing the set of trees calculated from the data.

```

1: set the AvailableClusters to all the elements in  $X$ .
2: set Edges  $\leftarrow \emptyset$ .
3: set Nodes  $\leftarrow$  AvailableClusters.
4: while there is a pair of clusters that can be joined do
5:   select  $P, Q$  such that  $d(P, Q) = \min_{P', Q' \in \text{AvailableClusters}} d(P', Q')$ .
6:   remove  $p, q$  from AvailableClusters and add  $P \cup Q$ .
7:   add  $P \cup Q$  to the Nodes of the output graph.
8:   add  $(P, P + Q), (Q, P + Q)$  to the Edges of the output graph.
9: end while
10: return the graph (Nodes, Edges).

```

major assets of hierarchical algorithms. These properties lead to the interactive exploration of the results as the user can interactively analyse the dendrogram and different thresholds resulting in different clusterings results. Furthermore, it is possible to design an objective function to automatically analyse the dendrogram and obtain a clustering which agrees with the desired properties. Hierarchical clustering algorithms are not only able to provide a flat clustering of the elements but can also obtain a hierarchy from them, allowing for a rich set of post-clustering analysis to be performed.

3.2.2 Partitional algorithms

Partitional algorithms result in a single partition of the data. In large sets, where the construction of the dendrogram is costly, partitional algorithms perform better. However, a common problem with partitional algorithms is that the number of clusters has to be determined in advance. In subsection 3.1.3, we referenced a number of heuristics designed to solve this problem, there is, however, no silver bullet. The partitional algorithms usually optimize a given criterion in order to perform the clustering. This criterion is defined either locally (on a subset of patterns) or globally (on the entire dataset). A combinatorial search of the optimal value for all possible labellings is clearly prohibitive. In practice, the algorithm converges to a local optimum and either provides this result as the clustering or repeats the operation in a randomized way to reach a locally optimal labelling which is closer to the global.

Squared-error Algorithms

The more frequently used algorithms minimize a squared error criterion, which works well with isolated and compact clusters. This partitional family of algorithms

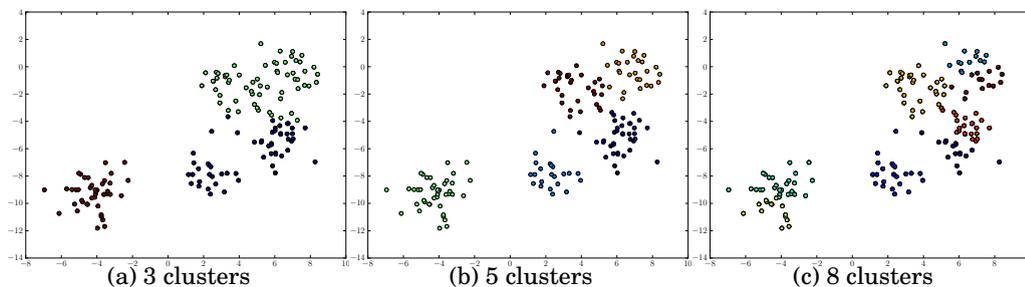


Figure 3.5: Results of the K-means algorithm using $d(x, x') = \|x - x'\|_2$ with several values of k . It can be observed that regardless the “real” number of clusters in the dataset, K-means always finds k clusters.

searches clusters which minimize a squared-error function, usually

$$\epsilon^2(C) = \sum_{i=1}^k \sum_{x \in C_i} d^2(\text{centroid}(C_i), x). \quad (3.1)$$

Each algorithm using $\epsilon(\cdot)$ as its objective function needs the definition of a cluster centroid. In the case of the most common squared-error algorithm, K-means, the centroid is the arithmetic mean:

$$\text{centroid}(C) = \underset{\bar{x}}{\text{argmin}} \frac{1}{|C|} \sum_{x \in C} d^2(\bar{x}, x).$$

The basic implementation of an algorithm of the K-means family is shown in algorithm 3.2 and an example of the application of K-means to a small dataset is illustrated in figure 3.5. One of the main drawbacks of most algorithms of this family is that, regardless of the “real” number of clusters in the dataset, they always find k clusters, putting into question the validity of their results.

Since Steinhaus [1956] proposed K-means, there have been several extensions to this algorithm. The most representative were designed in order to: find a better approximation to the optimal minima by adding criteria to split the sets of clusters [Ball and Hall: 1965, Forgy: 1965]; make it more robust to outliers by replacing the mean with the medoid, K-medoids [Kaufman and Rousseeuw: 1987]; provide a probability of belonging to the cluster instead of a hard partition, C-means [Dunn: 1973]; generalize it to arbitrary shapes, kernel K-means [Scholkopf et al.: 1998]; and to automatically find the number of clusters, X-means [Pelleg and Moore: 2000]. All of these variants have been used in different situations with reasonable success. However, combining them, for instance, in order to derive a version which is able to handle arbitrary shapes and automatically finds the number of clusters, is not simple. Regarding the inherent problem of error minimization in K-means, it has been shown that the minimization of its objective function (equation 3.1) is an NP-Hard problem [Meilă: 2006]. However, Meilă [2006] showed that when the clusters are well separated, it has a high probability of converging to the global optimum.

Algorithm 3.2 Squared-error clustering algorithm

Inputs: A dissimilarity function between elements, $d(\cdot, \cdot)$; a centroid function, $\text{centroid}(\cdot)$; a set of elements X ; and a number of clusters k .

Outputs: A set of clusters $\{C_1, \dots, C_k\}$ minimizing $\epsilon^2(C)$ (equation 3.1).

```

1: Pick partition of  $X$  in  $k$  clusters  $C = \{C_1, \dots, C_k\}$ 
2: repeat
3:   calculate the centroid of each cluster:  $c_i = \text{centroid}(C_i)$ 
4:   set  $C_{old}$  to  $C$ 
5:   generate a new clustering  $C = \{C_1, \dots, C_k\}$  s.t.
6:   for each element  $x \in X$  do
7:      $x \in C_i$  iff  $d(c_i, x) = \min_{j=1\dots k} d^2(c_j, x)$ 
8:   end for
9: until the new clustering  $C$  is equal to the old clustering  $C_{old}$ 
10: return  $C$ 

```

All in all, K-means is the most frequently used partitional algorithm. Despite all its limitations, its simple implementation and empirical results amount to the user's expectation in a wide variety of applications.

Mixture-resolving Algorithms

The mixture resolving approach assumes that the patterns to be clustered are drawn from one of several distributions, usually parametric, and the goal is to identify the parameters of these distributions. In some cases, the number of distributions is also inferred from the data. The previously analysed algorithm, K-means, can be formulated in terms of this paradigm as well. Its basic implementation amounts to infer the parameters of a fixed number of Gaussian distributions with different means and the same variance. The most well-known technique for mixture-resolving is the Expectation-Maximization (EM) algorithm [Dempster et al.: 1977] which finds the parameters of the mixture model by maximizing their likelihood with respect to the dataset. An advantage of this algorithm is that it assumes that the data represents a set of observed elements considering that unobserved elements also exist. Using this representation, the algorithm takes models of the observed and unobserved elements in order to obtain the parameters of the mixture. Using a Bayesian approach, the latent Dirichlet Allocation (LDA) algorithm proposed by Blei et al. [2003] and its derivatives [Blei et al.: 2004, Teh et al.: 2006] use an infinite mixture of Dirichlet processes in order add the number of clusters to the inferred parameters.

Mode-seeking Algorithms

Mode seeking is the general name for a family of clustering algorithms which assumes that the data to be clustered can be regarded as an empirical probability distribution function. In this function, dense regions correspond to the maxima of

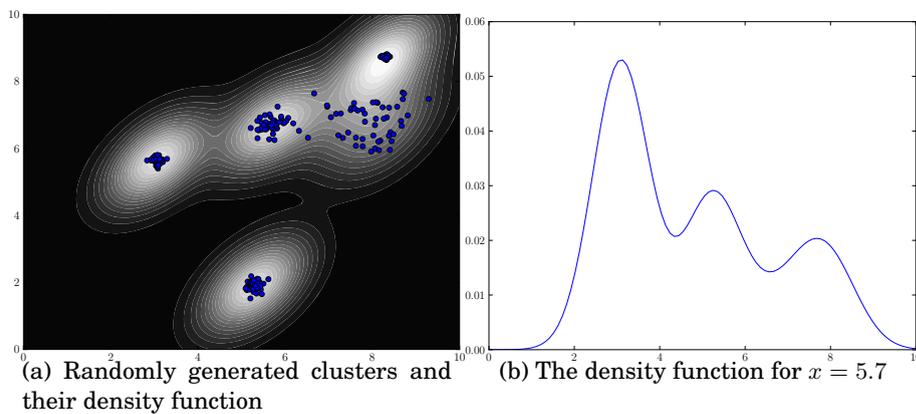


Figure 3.6: A set of randomly generated points and their estimated density function. Lighter colour in the background indicates higher density.

the p.d.f., usually called modes, representing cluster centroids. This concept is illustrated in figure 3.6 where a set of random generated cluster in 2D and their density function is shown. Mode seeking algorithms are designed in order to have two main features: there is no shape bias and there is no pre-specified number of clusters. However, they suffer from their main hypothesis, in high-dimensional spaces a region with high density is difficult to find making them ineffective for this case.

Generally speaking, these techniques search for densely connected regions in feature space called modes. Different algorithms use different definitions of connectedness. The Jarvis-Patrick algorithm [Jarvis and Patrick: 1973] defines that two points to cluster are as similar as the number of neighbours they share within a specified radius. The DBSCAN [Ester et al.: 1996] and mean shift algorithms [Cheng: 1995] use Parzen windows in order estimate the density of the points in space and then make their points go “up-hill” towards peaks in this density function. The main difference between DBSCAN and mean shift is their original goal. DBSCAN was conceived to perform clustering in large databases, hence more attention has been paid to its memory consumption and algorithmic cost. Mean shift, an application of the density gradient estimation algorithm of Fukunaga and Hostetler [1975], is more concerned with the mathematical properties of the density estimation and their consequences on the clustering result than with algorithmic costs. The main problem of the mean shift approach is that it looks for the clusters by optimizing a function for each element leading to convergence problems. Hinneburg and Keim [1998], without knowing mean shift clustering algorithm, generalized DBSCAN into DENCLUE using the density estimation proposed by Fukunaga and Hostetler [1975] but with more attention to algorithmic costs than Cheng [1995]. The mean shift algorithm finally made its entrance in the image processing area through the work of Comaniciu and Meer [2002] and Carreira-Perpiñán [2000]. Since then, it sprouted a copious amount of variations and applications having more than 2,000 citations eight years after its publication.

Algorithm 3.3 Mean shift algorithm

Inputs: A dissimilarity function between clusters, $d(\cdot, \cdot)$, a kernel function $\phi_h(\cdot)$ and a set of elements X .

Outputs: A set of clusters C .

```

1:  $C \leftarrow \emptyset$  //Set of resulting clusters
2: for each point  $\mathbf{x}_0$  in  $X$  do
3:   set  $\mathbf{x}_0 \leftarrow \mathbf{x}$ 
4:   set  $k \leftarrow 0$ 
5:   repeat //Look for the closest mode to  $\mathbf{x}$ 
6:     set  $\mathbf{x}_{k+1} \leftarrow \arg \min_{\mathbf{x}'} \sum_{\mathbf{x}' \in X} d^2(\mathbf{x}', \mathbf{x}) \phi_h(d^2(\mathbf{x}', \mathbf{x}_k))$ 
7:     set  $k \leftarrow k + 1$ 
8:   until  $\mathbf{x}_k = \mathbf{x}_{k-1}$ 
9:   if there is  $c \in C$  s.t. its mode is  $\mathbf{x}_k$  then
10:    add  $\mathbf{x}$  to  $c$ 
11:   else
12:    add the set  $\{\mathbf{x}\}$  to  $C$ 
13:   end if
14: end for
15: return  $C$ 

```

Mean shift uses a kernel function Φ in order to perform density estimation, given a distance function $d(\cdot, \cdot)$ (which is a dissimilarity function with a triangular inequality property). Using this kernel function, the empirical density for a point $\mathbf{x} \in X$ is given by

$$f(\mathbf{x}) = c \sum_{\mathbf{x}' \in X} \Phi_h(d^2(\mathbf{x}, \mathbf{x}')). \quad (3.2)$$

In this equation, c is calculated such that $f(\cdot)$ integrates to 1 and h is the bandwidth of the kernel function Φ . Having this density estimation, each point is moved towards its sample mean by following the gradient direction of the density estimation. Using equation 3.2 to estimate the local density of \mathbf{x} , and by notating $\mathbf{x}_0 = \mathbf{x}$, at each step of the algorithm the point is moved to a new location

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} \sum_{\mathbf{x}' \in X} d^2(\mathbf{x}', \mathbf{x}) \phi_h(d^2(\mathbf{x}', \mathbf{x}_k)). \quad (3.3)$$

where $\phi_h(\mathbf{x}) = \nabla \Phi_h(\mathbf{x})$. Finally, all the points that converged to the same mean are considered to be in the same cluster. The pseudocode of mean shift is shown in algorithm 3.3 and the behaviour of the algorithm is illustrated in figure 3.7. The arbitrary shape of the clusters can be noticed in this figure.

The mean shift algorithm has two disadvantages: convergence, the cycle starting in line 5 of algorithm 3.3 can oscillate indefinitely; speed, the algorithm has to minimize a cost function (equation 3.3) for each element to cluster. The speed problem has been addressed by using a variable step size for the cost function [Hinneburg and Gabriel: 2007] and by casting the algorithm in a EM framework and using this interpretation to achieve an optimization of the algorithm [Carreira-Perpinan: 2007].

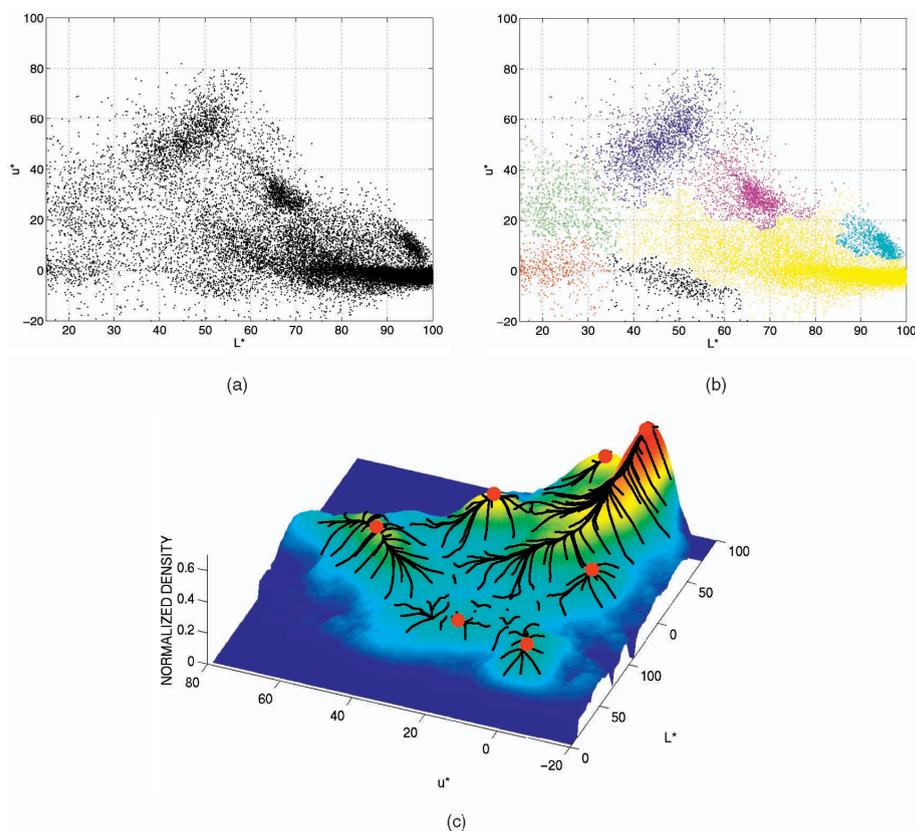


Figure 3.7: Illustration of mean shift clustering. (a) two dimensional set of 110,400 points. (b) Clustering result. (c) Trajectories of the mean shift procedures drawn over the density estimate of the dataset. Reproduced from [Comaniciu and Meer \[2002\]](#).

Finally, [Sheikh et al. \[2007\]](#) recently solved both issues by using the medoid instead of the mean: being the medoid an element of the input dataset, the problem to solve for each element changes from an optimization problem to a selection problem, leading to a faster algorithm with proved convergence.

Graph-based Algorithms

Graph theoretic-clustering represents the data points as nodes in a weighted graph and the edges connecting the nodes are weighted by their pair-wise similarity $s(\cdot, \cdot)$. The main goal of these algorithms is to find a partition of the nodes into two subsets A and B such that the similarity between the two sets is minimal. More formally: The input of the algorithm is a graph $G(X, E)$ where $X = \{x_1, \dots, x_N\}$ is the set of nodes and $E \subset X \times X$ the set of edges. Then, the algorithm must find two sets of nodes A and B such that $A \cap B = \emptyset$ and the inter-set similarity, or cut, defined as

$$\text{cut}(A, B) = \sum_{x_a \in A, x_b \in B} s(x_a, x_b) \quad (3.4)$$

is minimal.

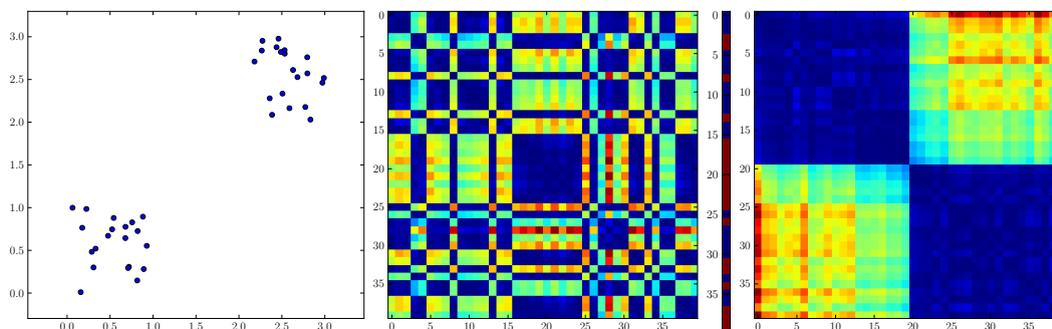


Figure 3.8: Illustration of a simple application of Fiedler's theory to clustering. From left to right: toy example to cluster; distance for the example matrix, where blue means that the elements are close while red means that they are distant; Fiedler vector for the graph induced by the previous matrix, blue means a value smaller than 0 and red larger; and the reordered distance matrix where the rows and columns are grouped according to their value on Fiedler's vector, the cluster structure is clearly observable.

The theoretical background for these algorithms was stated by Fiedler [Fiedler: 1973, 1975]. He, analysed the connectivity of graphs through the eigendecomposition of a matrix which was later called Graph Laplacian [Pothen et al.: 1990, Mohar: 1991, Chung: 1997] and is defined as

$$L(G)_{ij} = \begin{cases} \deg(\mathbf{x}_i) & i = j \\ -1 & (x_i, x_j) \in E \\ 0 & \text{any other case} \end{cases} . \quad (3.5)$$

where $\deg(\mathbf{x}_i)$ is the number of nodes connected to \mathbf{x}_i or its degree. [Fiedler: 1975] proved that the second smallest eigenvalue of this matrix is 0 if and only if G is not connected. Furthermore, he demonstrated that it is possible to use its corresponding eigenvector, known as the Fiedler eigenvector, to split the graph into two highly connected subgraphs while minimizing the cut size: taking advantage of the fact that each component of the Fiedler eigenvector corresponds to a node, he showed that clustering is simply performed by grouping the nodes whose value in this eigenvector is positive on one side and the ones with negative value on the other. A toy example of an application of this theory is shown in figure 3.8.

His work characterizing the connectivity of a graph through the eigenvalues of its Laplacian matrix was continued later as *Spectral Graph Theory*. In this theory, the notion of Graph Laplacian is generalized by allowing magnitudes different from 1 in the non-zero elements of the matrix and leading to multiple applications [Chung: 1997]. Furthermore, Hein et al. [2007] formalized the relationship between graph Laplacian and the Laplacian operator and studied in depth its consequences.

The main problem with the application of Fiedler's technique, known as Min-Cut clustering, is that the cluster sizes are unbalanced: if there is a very small set of isolated points, the output of the algorithm would be a very small and a huge cluster. This type of result does not shed any light on the structure of the clustered elements

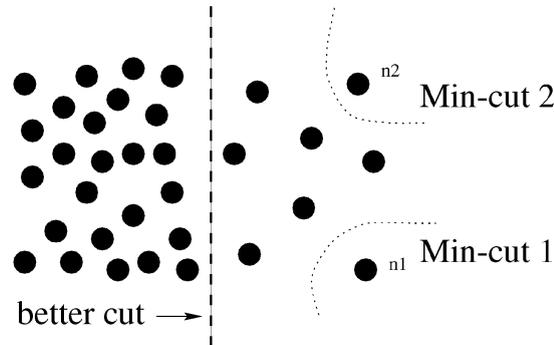


Figure 3.9: Unbalanced clustering using graph-theoretic techniques and the min-cut. Reproduced from [Shi and Malik \[2000\]](#)

nor it agrees with the goals of a clustering analysis, see figure 3.9. It was [Wei and Cheng \[1991\]](#) who proposed to solve this problem by normalizing the cut with the cluster sizes and named it *ratio-cut*. Later, [Hagen and Kahng \[1992\]](#) showed that the ratio-cut could be formulated into a Graph Laplacian and used the tools proposed by [Fiedler \[1975\]](#) to find the graph partition.

Not ten years after this, [Shi and Malik \[2000\]](#) boosted the use of these techniques. Firstly, they proposed the normalized cut,

$$NCut(A, B) = \frac{\sum_{x_a \in A, x_b \in B} s(x_a, x_b)}{\sum_{x_a \in A, x \in X} s(x_a, x) + \sum_{x_b \in B, x \in X} s(x_b, x)} \quad (3.6)$$

and showed that it was possible to formulate this problem into a Fiedler problem. In their work [Shi and Malik \[2000\]](#) prove that the normalized cut is a particular type of Graph Laplacian and that the cut is obtained by finding the median of the Fiedler eigenvector and splitting it in the values which are smaller and larger than its median. Moreover, they applied this algorithm recursively in order to obtain an arbitrary number of clusters.

More efficient extensions of this algorithm to handle an arbitrary number of clusters were proposed by [Meilă and Shi \[2001\]](#) and [Yu and Shi \[2003\]](#). [Ng et al. \[2002\]](#) improved it in order to handle clusters at different scales. Spectral clustering quickly became a fertile area with many variants and applications, for a survey on these techniques we direct the reader to [von Luxburg et al. \[2005\]](#) and [Filippone et al. \[2008\]](#). However, there is a main problem with this kind of clustering, a large eigenproblem must be solved, even if this problem has been addressed [[Fowlkes et al.: 2004](#)] the proposed solutions add parameters to the algorithm which have no interpretation regarding the clustering technique but a noticeable effect in the results.

Other Algorithms

More clustering algorithms based on different paradigms exists and have not been included in this survey. Just to name a few: [Tishby and Slonim \[2000\]](#) used infor-

mation theoretical methods, [Souvenir and Pless \[2005\]](#), [Haro et al. \[2006\]](#) topology and distribution mixtures. A more extensive review of clustering algorithms can be found in [Jain \[2009\]](#) and [Duda et al. \[2001\]](#).

3.3 SUMMARY

In spite of the large number of clustering applications and their success on different application domains, clustering remains a difficult problem. The vagueness of the definition of a cluster, changing from one application domain to the other, seems to be the main cause of this issue. Hence, it is the task of the user to analyse the dataset, to define the goal to achieve through clustering and to identify the reasonable hypotheses that can be formulated on the application domain. The result of this analysis should lead to a similarity or dissimilarity measure and to the selection, or invention, of a clustering algorithm.

In the particular case of this thesis, we use clustering techniques to develop algorithms which perform automated *in vivo* dissection of the brain's white matter structures. For this, clustering is applied to two types of objects: the voxels of an image acquired through DMRI [[Ziyan et al.: 2006](#), [Wiegell et al.: 1999](#), [Behrens et al.: 2003](#), [Lenglet et al.: 2006a](#), [Wassermann et al.: 2008](#)] and the fibres obtained using tractography algorithms [[O'Donnell and Westin: 2007](#), [Maddah et al.: 2008a](#), [Wassermann et al.: 2010a](#)]. The main issues to solve in these two clustering scenarios are which features to use for each type of element, how to represent them and how to measure the proximity between them. Once all of these issues have been solved, it is possible to choose an appropriate clustering algorithm. In the case of voxel clustering, the diffusivity profile within each voxel and its position must be taken into account. In the case of the clustering of white matter fibres, the main features to take into account are the shape, position and trajectory of a fibre. Further, in some cases it is useful to take into account the diffusivity information along them. The two subsequent parts of this thesis analyse these scenarios and develop clustering algorithms for DMRI image voxels and fibres obtained from tractography. Consequently, each one of the following parts of this thesis leads to the automatic identification of white matter structures in the brain.

Part II

White Matter Structure Identification Through Manifold Learning

Overview

Part II covers a new clustering algorithm which performs *in vivo* dissection of white matter structures by clustering the voxels of a DMRI image. In presenting this algorithm, we study several characteristics of the elements to cluster and of the expected clustering output. First, we analyze which diffusivity model, DT or ODF, is more appropriate for this task and how to measure similarity of the diffusivity profile among two voxels in both cases. Then, we study how spatial location of the voxels and their diffusivity information can be integrated in a single similarity measure. Next, we analyse the available hypotheses about the shape, linear separability and quantity of the clusters we are expecting to find. Due to the lack of assumptions about the shape and separability of the clusters that is possible to make, we propose the use of manifold learning techniques in order to find an appropriate representation of the elements to cluster. Particularly, we advocate the use of the Diffusion Maps manifold learning technique which requires a smaller number of hypotheses from the data to be clustered than current approaches in the literature. Finally, we use the previous analysis in order to develop an algorithm to cluster the voxels of a DMRI image into white matter structures and we validate this algorithm in synthetic and real data. The contributions presented in this part have been published in [Wassermann et al. \[2008\]](#) and [Wassermann et al. \[2007\]](#).

Contributions of this part:

- Definition of a similarity measure for DT and QBall image voxels which combines spatial and diffusivity information at each voxel.
- Development of a new Diffusion Maps-based clustering algorithm which requires a smaller number of hypotheses on the data than current algorithms in the literature.
- Analysis of the behaviour of this algorithm in synthetic and real data images. Comparing the results of its DT and QBall versions.

Introduction

Is it possible to automatically group the voxels of a diffusion magnetic resonance image such that the results of this grouping are the white matter structures? In this part of the thesis, we address the problem of *in vivo* dissection of the white matter by clustering the voxels of DMRI images, particularly, in the case of DT and QBall images. In order to develop a DMRI voxel clustering algorithm, three analyses should be performed, as seen in chapter 3. Firstly, we need to identify which features of the voxel should be considered. Next, we must select a proximity measure between the voxels using these features. Finally, we should analyse which hypotheses we can state about the clusters we expect as an outcome. The development of an appropriate clustering algorithm to identify white matter structures in DMRI images comes from the detailed analysis of these problems.

In the context of DMRI image clustering, two main features identify each voxel: the position on the image and the diffusion characteristics. This highlights the need of a proximity measure between diffusion tensors (DTs) and orientation distribution functions (ODFs). Examples of such measures have been introduced in chapter 2. In terms of hypotheses about the clustering outcome, it is difficult to state an assumption about the quantity or shape of the clusters. We can not affirm that the clusters we are expecting to find are convex or linearly separable. However, this can be solved by using manifold learning techniques, introduced in chapter 4, in order to infer a representation of the elements in which the clusters are convex and linearly separable. Once an appropriate representation for the voxels has been inferred, it is feasible to apply a wide range of clustering algorithms in order to identify white matter structures. In the current part of this thesis, we take into account the previously stated analysis and develop a clustering algorithm for white matter structure identification in DMRI images.

We start by reviewing recent literature on DMRI voxel clustering and then stating our contributions in this area.

State of the Art on DMRI Voxel Clustering

Clustering methods for diffusion MRI have been recently introduced. They typically rely on some metric between DTs or ODFs and allow us to identify various fibre bundles or regions of the white matter with different diffusion profiles. While

many techniques have been proposed to classify the grey matter, white matter and cerebrospinal fluid from T1-weighted MR images, the literature addressing the clustering of white matter and sub-cortical structures from diffusion MRI is fairly recent. To perform effective clustering, the contribution of the diffusivity and spatial location features must be carefully exploited. Particularly, the quantification of the similarity between the diffusion features (tensors, ODFs) is still a subject of current investigations. In the following we survey recent clustering methods for DMRI images. For this, we denote the position in the diffusion image by \mathbf{x}_i and \mathcal{D}_i stands for the diffusion characteristic (either the diffusion tensor or some representation of the ODF).

Methods Based on DTI

The first approach that used DTI to elucidate structure in the brain by means of clustering was designed to identify the different nuclei of the thalamus by [Wiegell et al. \[2003\]](#). It uses a k -means algorithm. The spatial metric is the Mahalanobis distance with respect to each cluster and the feature metric is the Frobenius norm of the difference between tensors. The choice of this last metric is crucial and discussed in the following, where we focus on fibre bundle segmentation.

[Zhukov et al. \[2003\]](#) proposed one of the very first approaches to fibre bundle segmentation, in this approach he was able to cluster white matter structures by only using the fractional anisotropy as the diffusion characteristic, in a surface evolution framework (which is well-suited for controlling the shape and smoothness of the resulting clusters). A 3D surface \mathcal{S} is represented by the zero level set of a 4-dimensional function ϕ ,

$$\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^3 : \phi(\mathbf{x}, t) = 0\},$$

and ϕ is evolved according to the differential equation

$$\partial\phi(\mathbf{x}, t)/\partial t = -F(\mathbf{x})\|\nabla\phi(\mathbf{x}, t)\|,$$

where F is a scalar-valued function which drives the evolution of ϕ , and implicitly deforms the surface \mathcal{S} along its normals. F is usually made of two terms

$$F = F_c + \beta F_s.$$

F_c quantifies characteristics of the regions to segment and F_s drives the smoothness of the surface; β is a user-selected weight. F_s and F_c can be respectively chosen as the mean curvature of the surface \mathcal{S} and an edge detector function $\mathbb{R} \mapsto [0, 1]$ applied to a smoothed FA map [[Zhukov et al.: 2003](#)]. To generalize this approach and take advantage of the full tensor information, [[Feddern et al.: 2003](#)] adapted the function F_c to work on a generalized structure tensor for diffusion tensor fields. This approach allowed some improvements over the previous work and was among the first to fo-

cus on the definition of an adequate metric between diffusion tensors. Later, [Wang and Vemuri \[2004\]](#) introduced a statistical Level-Set segmentation method. In this method, F_c is based on a regional description of the inner and outer compartments:

$$F_c = -d_f(\mathcal{D}, \bar{\mathcal{D}}_{in}) + d_f(\mathcal{D}, \bar{\mathcal{D}}_{out}) = -\|\mathcal{D} - \bar{\mathcal{D}}_{in}\|_F^2 + \|\mathcal{D} - \bar{\mathcal{D}}_{out}\|_F^2, \quad (3.7)$$

where $\bar{\mathcal{D}}_{in}$ and $\bar{\mathcal{D}}_{out}$ are the Fréchet means of tensors inside and outside the surface S . The Fréchet mean $\bar{\mathcal{D}}$ of a set of N tensors $\{\mathcal{D}_i\}_{i=1,\dots,N}$ is analytically or iteratively computed as the minimizer of the tensors' variance, depending on the choice of metric. Such a regional approach allows the tensors in the inner and outer regions to vary in a piecewise constant manner, contrary to approaches which only search for sharp variations of the FA, or other anisotropy maps [\[Zhukov et al.: 2003\]](#). Thus, the algorithm presented by [Wang and Vemuri \[2004\]](#) is capable of detecting fibre bundles where the tensor's shape changes smoothly. However, because of the use of the Euclidean distance d_f between tensors (Frobenius norm), the Fréchet mean $\bar{\mathcal{D}}_{in/out}$ is not guaranteed to be positive-definite, thus generating artefacts and incorrect segmentations in regions where tensors' variation is large. To overcome this problem, several authors have studied the influence of the metric. Considering the diffusion tensor as the covariance matrix of a zero-mean Gaussian distribution, [Wang and Vemuri \[2005\]](#) proposed the symmetrized Kullback-Leibler divergence or J-Divergence:

$$d_j(\mathcal{D}_1, \mathcal{D}_2) = \frac{1}{2} \sqrt{\text{trace}(\mathcal{D}_1^{-1} \mathcal{D}_2 + \mathcal{D}_1 \mathcal{D}_2^{-1}) - 6},$$

and used to extend previous work [\[Wang and Vemuri: 2004\]](#). This is a natural metric between probability distributions, which turns out to have a closed form expression in the Gaussian case. It also has a closed form expression for the mean tensor. Next, [Jonasson et al. \[2005\]](#) proposed two different distances between tensors in a similar Level-Set formulation. This was later extended by [Jonasson et al. \[2007\]](#) to prevent overlapping when propagating multiple surfaces for the segmentation, for instance, of the thalamic nuclei. The first distance, called Integrated Similarity, compares the diffusion properties from two different voxels. It is expressed as

$$d_{is}(\mathcal{D}_1, \mathcal{D}_2) = \frac{1}{4\pi} \int_{S^2} \min\left(\frac{\mathbf{g}_1(r)}{\mathbf{g}_2(r)}, \frac{\mathbf{g}_2(r)}{\mathbf{g}_1(r)}\right) dr,$$

where $\mathbf{g}_1(r)$ is the diffusion coefficient in direction r for the tensor \mathcal{D}_1 . This metric compares diffusion coefficients over all possible directions and is very sensitive to small differences between the shapes of the tensors. It has, however, a high computational cost. Another metric is used to calculate the empirical mean of a set of tensors. It measures the overlap between two tensors and is defined as

$$d_o(\mathcal{D}_1, \mathcal{D}_2) = \sqrt{\text{trace}(\mathcal{D}_1 \mathcal{D}_2)}.$$

Lenglet et al. [2006a] proposed to use a Riemannian metric derived from the Fisher information matrix in an extended statistical framework. This metric yields a geodesic distance on the manifold of zero-mean Gaussian distributions $S^+(3)$, as presented in chapter 2. This distance can also be expressed as

$$d_g(\mathcal{D}_1, \mathcal{D}_2) = \sqrt{\frac{1}{2} \sum_{i=1}^3 \log^2(\lambda_i)}$$

where the scalars λ_i are the eigenvalues of the matrix $\sqrt{\mathcal{D}_1^{-1} \mathcal{D}_2} \sqrt{\mathcal{D}_1^{-1}}$. The authors also demonstrated how to approximate a Gaussian distribution on $S^+(3)$ and to exploit this information in the segmentation procedure. To provide comparison amongst previously mentioned DT approaches, Lenglet et al. [2006a] illustrated the practical differences of using the three different distances, $d_{f/j/g}$, within his extended statistical surface evolution framework. Recently, Arsigny et al. [2006] introduced a distance with similar properties to those of d_g . This Log-Euclidean distance

$$d_{le}(\mathcal{D}_1, \mathcal{D}_2) = \sqrt{\text{trace}((\log(\mathcal{D}_1) - \log(\mathcal{D}_2))^2)}$$

has the advantage to be simple to implement and fast to compute. However, it has not been extensively studied for segmentation tasks yet, aside from its use in two recent papers from Weldeselassie and Hamarneh [2007] and Malcolm et al. [2007]. In both works, the image is segmented into two parts by minimizing an energy functional similar to the one of Wang and Vemuri [2005]. A non-parametric approach relying on the Log-Euclidean distance and a Markov random field framework was also recently described [Awate et al.: 2007]. Finally, Ziyang et al. [2006] used a graph-theoretical approach, known as N-Cuts [Shi and Malik: 2000]. This graph partitioning technique is based on the link between the second smallest eigenvector of the Laplacian matrix of a graph and optimal partitions. The nodes of the graph are the voxel \mathbf{x}_i of the image and the weights of the edges between those nodes are obtained from similarities between neighbouring tensors. The similarity between tensors can be any of the previously described distances or restricted to the directional information of the principal eigenvectors. The outline of the procedure proposed by Ziyang et al. [2006] is as follows: First, a matrix W_s is built to encode local similarity between tensors. It is only non-zero for neighbouring voxels:

$$\{W_s\}_{ij} = \begin{cases} \exp\left(-\frac{d^2(\mathcal{D}_i, \mathcal{D}_j)}{\sigma^2}\right), & \text{if } \|\mathbf{x}_i - \mathbf{x}_j\| \leq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (3.8)$$

σ is a chosen scale parameter. Next, local similarities are propagated to a full affinity matrix W by converting W_s into a one-step transition probability matrix whose rows and columns sum to one. Markovian relaxation [Tishby and Slonim: 2000] is

used to generate the n -step transition probability matrix. Finally, this matrix is recursively partitioned in two clusters by using eigenvalue decomposition and thresholding the eigenvector with the second smallest eigenvalue. This produces a hierarchical clustering. The number of recursions is ultimately chosen to obtain the desired number of clusters. The main limitation of this algorithm is the need for a uniform sampling of the $(\mathbf{x}_i, \mathcal{D}_i)$ at the end of the Markovian relaxation as it will become more evident after we expose the basis of manifold learning in chapter 4. As we are now going to discuss, level set, Markov random fields and graph-theoretic segmentation frameworks have also recently been extended to HARDI datasets.

Methods based on HARDI

As stated in Section 2, the diffusion tensor model cannot describe complex white matter fibre configurations, and HARDI techniques like QBI were introduced to overcome this issue. It is thus natural to exploit this information to improve white matter segmentation results. The 5D space defined by the location of the ODFs on the acquisition grid and their orientational information can be used [Hagmann et al.: 2006, Jonasson et al.: 2006]. These segmentation procedures are respectively implemented with a hidden Markov random field or level set framework. McGraw et al. [2006] proposed to use Mixtures of von Mises-Fisher distributions to model the ODFs and perform segmentation with a hidden Markov random field scheme in their work. It is possible to use a spherical harmonics (SH) decomposition of the ODF at each voxel \mathbf{x}_i . The diffusion characteristic \mathcal{D}_i is then replaced by a vector of SH coefficients. Although the most appropriate metric between ODFs is an open area of research, the L_2 norm can be used and efficiently computed. Two other clustering techniques have been proposed that take advantage of the SH representation. First, Descoteaux and Deriche [2008], generalized the level set algorithm presented by Wang and Vemuri [2004] and Lenglet et al. [2006a] to the HARDI case. Equation (3.7) is modified such that the distance is the L_2 norm of the difference of SH coefficients.

There are common problems to all of these HARDI-based methods. Firstly, all of these methods are based on Level-Set techniques [Cremers et al.: 2007]. These techniques are designed to divide the image in two regions, a main object and the background. When we reviewed the anatomy of the white matter in chapter 2, we have seen that there is a large quantity of white matter bundles and that identification of each one of them on a one-by-one basis might be far away from the users goals.

Secondly, these techniques work as Gaussian-mixture separation techniques (chapter 3). They assume that the two regions that will be separated can be modelled as Gaussian distributions and analyse the image in order to find a partition that optimizes the parameters of these distributions. This mixture separation approach is not appropriate in the general case of white matter tract segmentation. The main

reason of this is due to the fact that the previously described algorithms use diffusion models on each voxel based on the directionality of diffusion. In bundles like the cortico-spinal tract this seems appropriate. The diffusion along this tract is mostly oriented in the superio-inferior direction, hence the distribution of the diffusion directionality is distinctive. However, if we analyse a C-shaped tract, like the arcuate or the frontal forceps, the diffusion profile of the voxels within this tracts takes, at least, all the directions within a plane making it difficult to separate this tract from the rest through mixture-separation.

Our Contribution

In order to improve the previous techniques, we propose a new method that can segment fibre bundles using QBall images as an input. Moreover, our algorithm deals with fibre crossings while requiring a minimum number of hypotheses from the data and a small number of algorithmic parameters.

The first improvement with respect to the HARDI-based techniques that have been reviewed is the fact that our algorithm is prepared to handle an arbitrary number of clusters and requires no manual initialization. This is achieved by using the K-means algorithm presented in chapter 3. In this particular application, the implementation of this algorithm is far from being trivial, in fact it has to deal with two major problems.

The choice of a proximity measure between ODFs is the first problem to tackle. Even if the proximity measure presented for this purpose in chapter 2 is general for all the functions representable by spherical harmonics [Descoteaux et al.: 2007a], it is not suited for clustering. While these functions are in a high-dimensional space, the ODFs, a subset of these, are in a subspace which has a smaller number of dimensions. In order to infer a proper space for the ODFs along with a more appropriate metric to quantify their similarity, we use manifold learning techniques.

Once we have found a proximity measure that is adjusted to our problem, we must deal with feature integration. In this application, K-means clustering is expected to group voxels of the QBall image into white matter fibre bundles. Hence, the algorithm must handle is the integration of the diffusion information inside each voxel and the location of the voxel in space.

In the following chapters, we present the theoretical basis that our algorithm uses in order to solve these two problems: the first issue is the modelling of the manifold which contains the ODFs of the white matter. In order to solve this problem we use manifold learning techniques. These techniques take as an input the elements of the manifold we want to cluster and a general proximity measure and *learn* a proximity measure which renders the clustering problem simpler. Then, we describe how we integrate information about the position of the voxel in space and its diffusivity profile through markovian relaxation.

The two following chapters are as follows: First we review the theoretical basis

of manifold learning in general and of the manifold technique (Diffusion Maps) we chose to solve our problem . In the subsequent chapter, we describe our algorithm, the markovian relaxation and we show results on synthetic images and real data.

MANIFOLD LEARNING

We have seen in subsection 3.1.1 that one of the main problems in clustering is the representation of the elements. In several applications these elements are either in a high dimensional space or in a sub-manifold of the space from which they were sampled. Take for instance the spiral set presented in figure 4.1a. In this set we can see points drawn from three Gaussian distributions with different parameters. Each of these distributions has generated a cluster, which should produce a high-density area in element space. However, the Gaussians are embedded on a spiral and the isolevels of the estimated density show that there is a high density area around the centre of the spiral and that it decreases towards the edges of the image. Three areas with distinctive density can be spotted, one in the centre and two on the lower right-hand side. None of these areas corresponds to the expected clustering result. However, if we unroll the spiral and estimate the density on this one-dimensional function (figure 4.1b), it clearly shows the three clusters as expected.

This example highlights the need of working in the right space, or manifold, for the data. However, it is not always possible to know in advance in which manifold the objects to cluster lie. This difficulty leads to a set of techniques which aim to automatically learn the manifold: *manifold learning techniques*.

4.1 FEATURE MATRIX-BASED TECHNIQUES

4.1.1 Principal Component Analysis

Pearson [1901] proposed the first manifold learning technique, Principal Components Analysis (PCA). In the original publication, Pearson stated that “In many physical, statistical and biological investigations it is desirable to represent a system of points in plane, three or higher dimensioned space by the best fitting straight line or plane”. Consequently, the goal of PCA is to find a projection of the dataset into an orthogonal coordinate system. The first coordinate of this system accounts for the direction of largest variance, the second for the second largest variance direction and so on.

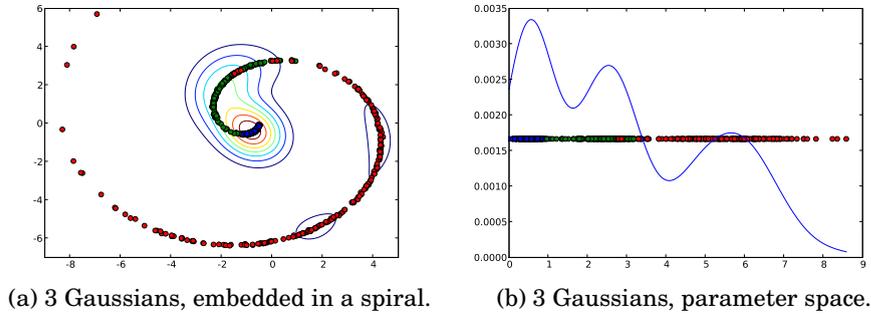


Figure 4.1: Importance of working in the right manifold for clustering. Given 3 clusters embedded in a curve. The clusters have different Gaussian distribution. On the left: density estimation on the bi-dimensional space enclosing the curve are shown as contour plots. Concentration areas do not coincide with clusters. On the right kernel density estimation along the curve: peaks coincide with the clusters.

It is possible to state the PCA algorithm as a maximization problem. The PCA algorithm takes as an input a set of N elements, more particularly random vectors $\mathbf{x} \in \mathbb{R}^D$, the elements of the set must be represented as $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$, its transposed feature matrix. Assuming that these elements are centred, $\sum \mathbf{x} = \mathbf{0}$, that $N \gg D$, and that $U = \{\mathbf{u}_1, \dots, \mathbf{u}_D\}$ is an orthogonal set of vectors. The goal of PCA is to find a projection of \mathbf{X} into U ,

$$\mathcal{P}(\mathbf{X}, U) = \sum_{k=1}^D \mathbf{u}_k \langle \mathbf{u}_k, \mathbf{X} \rangle, \quad (4.1)$$

such that \mathbf{u}_1 is aligned with the direction of maximal variance, \mathbf{u}_2 to the direction of largest variance orthogonal to \mathbf{u}_1 , and the other orthogonal directions, like the first two, are chosen in decreasing order of variance.

We start by calculating \mathbf{u}_1 . Due to the previous hypotheses, the linear combination $x = \mathbf{X}^T \mathbf{u}_1$ must possess maximum variance. Then, finding the direction \mathbf{u}_1 is equivalent to calculate the maximum of

$$\rho = \mathbb{E}[xx] = E[\mathbf{u}_1^T \mathbf{X} \mathbf{X}^T \mathbf{u}_1] = \mathbf{u}_1^T \mathbb{E}[\mathbf{X} \mathbf{X}^T] \mathbf{u}_1 = \frac{\mathbf{u}_1^T \mathbb{E}[\mathbf{X} \mathbf{X}^T] \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1}$$

where $\mathbb{E}[\cdot]$ is the expected value. As \mathbf{X} is centred by hypothesis, $\mathbf{C} = \mathbb{E}[\mathbf{X} \mathbf{X}^T] = \mathbf{X} \mathbf{X}^T$ is positive semi-definite and is the covariance matrix of \mathbf{X} . Then, the problem is reduced to finding the maximum of

$$\rho = \frac{\mathbf{u}_1^T \mathbf{C} \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1}.$$

This can be done by considering the gradient of ρ ,

$$\frac{\partial \rho}{\partial \mathbf{u}_1} = \frac{2}{\mathbf{u}_1^T \mathbf{u}_1} (\mathbf{C} \mathbf{u}_1 - \rho) \mathbf{u}_1.$$

Setting the gradient to 0 gives

$$\mathbf{C} \mathbf{u}_1 = \rho \mathbf{u}_1$$

which is recognized as the eigenproblem and ρ is maximized when it is the largest eigenvalue of \mathbf{C} . Being positive semi-definite, \mathbf{C} admits an eigendecomposition

$$\mathbf{C} = \sum_{i=1}^N \lambda_i \mathbf{e}_i \mathbf{e}_i^T = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T = (\mathbf{V} \sqrt{\mathbf{\Lambda}}) (\mathbf{V} \sqrt{\mathbf{\Lambda}})^T$$

where the eigenvalues $\lambda_1 \geq \dots \geq \lambda_D$ are positive, their corresponding eigenvectors, $\mathbf{V} = [\mathbf{e}_1, \dots, \mathbf{e}_D]$, are orthonormal and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_D)$. Therefore, the problem of maximizing the variance is equivalent to calculating the largest eigenvalue of λ_1 , λ_1 , and its corresponding eigenvector since

$$\lambda_1 = \mathbf{e}_1^T \mathbf{C} \mathbf{e}_1 = \max_{\mathbf{u}_1} \frac{\mathbf{u}_1^T \mathbf{C} \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1} = \max \rho.$$

Then, the first eigenvector, \mathbf{e}_1 describes the data in the direction of its maximal variance. If we are looking for a d -dimensional representation, $d < D$, of the elements of X , we still need $d - 1$ directions. Remove the components of \mathbf{X} in the direction of \mathbf{e}_1 and calculating a new maximum variance direction, we find that the second eigenvector of \mathbf{C} , the one which corresponds to λ_2 , is the best choice. This is because it indicates a direction, orthogonal to \mathbf{e}_1 , in which there is maximal variance providing we previously removed the components of the data in the direction of \mathbf{e}_1 . Following this idea, the basis is given by the eigenvectors of \mathbf{C} which correspond to the first d largest eigenvalues.

The PCA coordinate system is given by the basis $U = \{\mathbf{e}_1, \dots, \mathbf{e}_D\}$ and the projection of a new element \mathbf{x} to this coordinate system is given by

$$\mathcal{P}(\mathbf{x}, U) = (\langle \mathbf{x}^T, \mathbf{e}_1 \rangle, \dots, \langle \mathbf{x}^T, \mathbf{e}_D \rangle). \quad (4.2)$$

Finally, the d -dimensional projection is given by

$$\mathcal{P}_d^{\text{PCA}} = (\langle \mathbf{x}^T, \mathbf{e}_1 \rangle, \dots, \langle \mathbf{x}^T, \mathbf{e}_d \rangle) \quad (4.3)$$

An example of a dataset and its projection on a coordinate system obtained through PCA is shown in figure 4.2. However, as shown in figure 4.3, when the manifold to learn is not linear, the coordinate system which PCA obtains does not reflect the implicit coordinate space of the dataset.

Principal Components Analysis performs very well when confronted with datasets

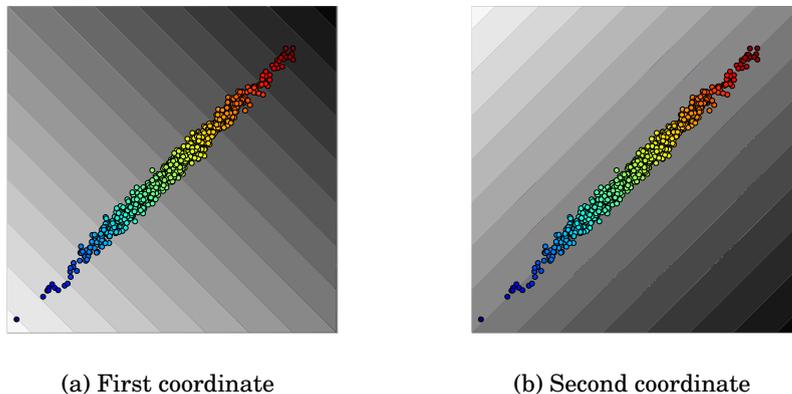


Figure 4.2: Principal Component Analysis coordinate on a linear dataset. The original dataset is shown as scattered points. The two directions of the coordinate system are illustrated by the direction of the grey-coloured degrade. It is noticeable how the coordinate system is aligned with the directions of maximum and minimum variability of dataset.

which can be projected into a hyperplane by squashing one or various of its dimensions. However, most of the datasets coming from observations of real data do not have this property. In fact, in a graphical representation where each element is a point in space, these datasets describe a non-linear shape. We illustrate this in figure 4.3 where the coordinate system for a “C”-shaped dataset is shown. In this figure, we observe how the orthogonal coordinate system does not describe the dataset at all. Scholkopf et al. [1999] proposed an algorithm to solve this problem which combines kernel methods with PCA (KPCA) extending PCA in order to extend it to non-linear scenarios. Still, PCA and its variants are designed in order to map elements from one coordinate system to another, based on projections. In the case of this thesis, the elements that we analyse, diffusion tensors, white matter fibres, etc., are usually not embedded in coordinated spaces, they are just comparable through proximity measures. For this reason, we focus our attention on manifold learning techniques apt for this scenario.

4.2 PROXIMITY MATRIX-BASED TECHNIQUES ---

If the dimensionality of the elements we analyse is large, PCA-based techniques will end up using huge matrices. Moreover, if the elements are not in \mathbb{R}^D , these methods are capable to deal with them. As this is usually the case in diffusion imaging, we direct our attention to methods which are based on proximity matrices.

4.2.1 Multi-Dimensional Scaling

Multi-Dimensional Scaling (MDS) was proposed by Torgerson [1952]. The main idea was to explain psychometric experiments on people’s perception of the similar-

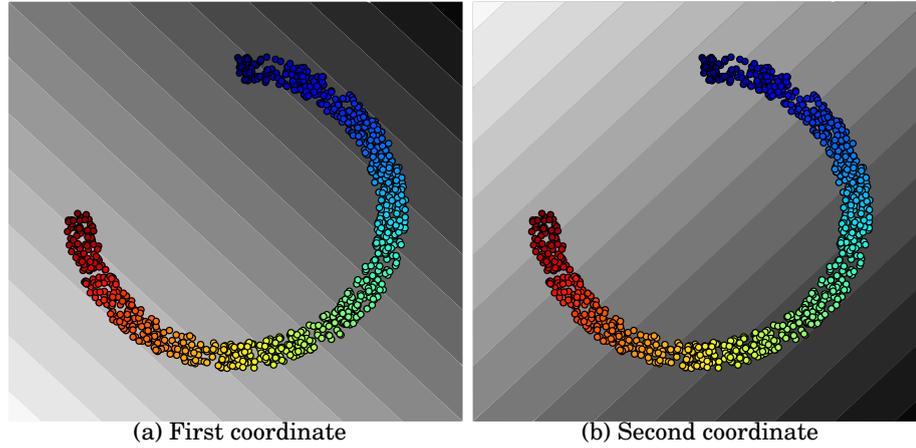


Figure 4.3: Principal Component Analysis coordinate on a non-linear dataset. The original dataset is shown as scattered points. The two directions of the coordinate system are illustrated by the direction of the grey-coloured degrade. It is noticeable how the orthogonal coordinate system is not adapted to the dataset.

ities inside classes of objects. The main advantage of MDS over PCA is that it is based on proximity matrices, hence the elements are inherently represented by a quantified degree of proximity instead of explicitly through features. The MDS algorithm maps a set of elements represented by its proximity matrix into a set of D -dimensional vectors in euclidean space.

Starting from a set of elements $X = \{x_1, \dots, x_N\}$ and a dissimilarity function $d(\cdot, \cdot)$, MDS builds a dissimilarity matrix $\mathbf{D}_{ij} = d^2(x_i, x_j)$. From \mathbf{D} , a similarity matrix \mathbf{S} is obtained by means of $\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T$, the centring matrix:

$$\mathbf{S} = \mathbf{H}\mathbf{D}\mathbf{H}. \quad (4.4)$$

In this case, \mathbf{S}_{ij} is proved to be an inner product whether $d(\cdot, \cdot)$ is a metric or not [Torgerson: 1952]. Then, as \mathbf{S} is symmetric positive semi-definite and of full-rank, there exists a squared, centred, matrix \mathbf{Y} such that $\mathbf{S} = \mathbf{Y}^T\mathbf{Y}$ with $\mathbf{S}_{ij} = \mathbf{Y}_i^T\mathbf{Y}_j$ [Torgerson: 1952]. Where \mathbf{Y} is a feature matrix such that \mathbf{y}_i is the result of the MDS mapping of \mathbf{x}_i to an N -dimensional space. Consequently, the spectral decomposition of \mathbf{S} can be used to infer \mathbf{Y} :

$$\begin{aligned} \mathbf{S} &= \mathbf{Y}^T\mathbf{Y} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \\ \mathbf{Y} &= \sqrt{\mathbf{\Lambda}}\mathbf{V}^T \end{aligned}$$

This matrix \mathbf{Y} is the transposed feature matrix providing a representation of the elements of \mathbf{X} in euclidean space. Finally, we build the d -dimensional mapping using the eigendecomposition of \mathbf{S} :

$$\mathcal{P}_d^{\text{MDS}}(\mathbf{x}_i) = \left(\lambda_1^{-\frac{1}{2}}(\mathbf{u}_1)_i, \dots, \lambda_d^{-\frac{1}{2}}(\mathbf{u}_d)_i \right)$$

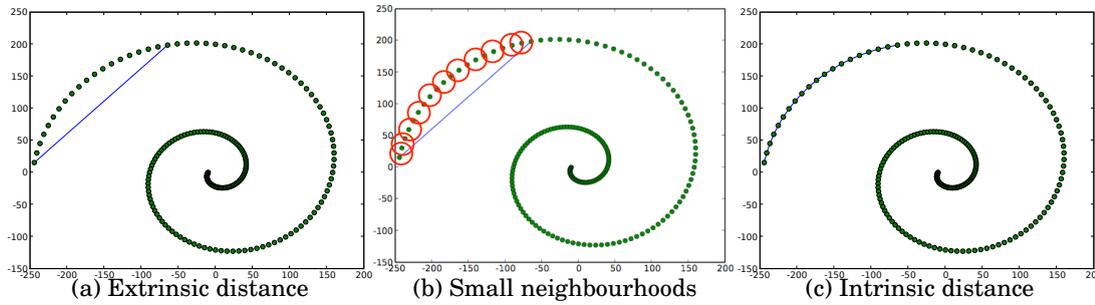


Figure 4.4: Isomap idea: subfigure (a) shows a dataset and its extrinsic distance as a blue line; subfigure (b) shows the previous dataset with red circles showing the limiting neighbourhoods used to compute the distance; and subfigure (c) shows how computing the distance through partial distances between nodes within the same neighbourhood produces the intrinsic distance.

The main advantage of MDS is that it produces an embedding into an euclidean space of any dataset, provided that we have a dissimilarity function or, by skipping the step in equation 4.4, an inner product function. Moreover, it can be proved that PCA and KPCA are special cases of MDS [Etynger: 2007].

4.2.2 Isomap

Perhaps the most successful specialization of MDS is Isomap. In order to perform non-linear manifold learning, Tenenbaum et al. [2000] proposed to use a pruned graph to compute the distance between elements. As shown in figure 4.4, by restricting the distance computations to small distances within the same neighbourhood, the intrinsic distance can be calculated. The Isomap algorithm is simple and effective in several scenarios.

Isomap starts with the distance matrix between every pair of elements, $D_{ij} = d(x_i, x_j)$, and then it prunes the matrix in one of two ways

ϵ -neighbours: D_{ij} is set to 0 if $D_{ij} > \epsilon$

k-neighbours: D_{ij} is set to 0 if $D_{ij} > D_{ii_k}$ where i_k is the k th closest element to i .

The k -neighbours technique does not ensure a symmetric matrix, hence D has to be symmetrized, for instance by setting $D_{ij} = \min(D_{ij}, D_{ji})$. Then, we take D as the connectivity matrix of a graph (where 0 means that the nodes are not connected) and Dijkstra's algorithm is executed in order to obtain the distance between each pair of nodes in the graph. The output of Dijkstra's algorithm, matrix D' , is fed to the MDS algorithm and the embedding is obtained.

The core point in this algorithm is the theorem proposed by Bernstein et al. [2000]. This theorem states that the distances obtained by Dijkstra's algorithm applied to the pruned graph converge asymptotically to the geodesic distances on the manifold described by the data, as the number of elements in the sample approaches infinity [Bernstein et al.: 2000]. An example of the Isomap algorithm is shown in figure 4.5.

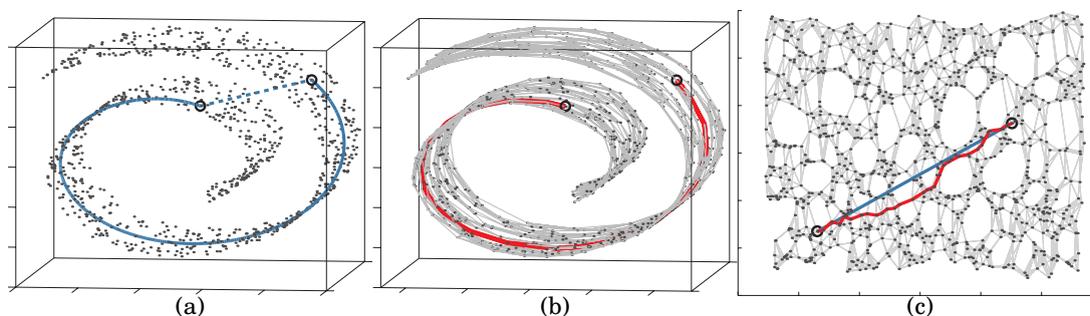


Figure 4.5: Example of Isomap embedding. From left to right: the input dataset where the intrinsic and extrinsic distances are shown; the pruned graph and the geodesic on this graph showing its similarity with the intrinsic distance; and the dataset embedded in two dimensions where it is possible to observe the edges of the pruned graph, the intrinsic distance and the estimated distance using Dijkstra's algorithm. Adapted from [Tenenbaum et al.: 2000]

4.2.3 Laplacian-based methods

The main idea behind these methods is to consider explicitly that the data lies on a low dimensional manifold embedded in a high dimensional space. Laplacian-based methods for dimensionality reduction, proposed by Belkin and Niyogi [2003], construct a map which embeds the original dataset into a low dimensional euclidean space. This mapping is built such that it preserves local neighbourhood topology and retains algorithmic simplicity. Instead of performing a global embedding which aims to preserve the distance relationship between every pair of elements in the manifold, these techniques maintain the local relationships only. Like the previous methods, the actual algorithms are reduced to perform certain operations on a similarity matrix and then use its eigendecomposition to construct the mapping. The main advantage of these methods is their sound mathematical basis in which the projection to the low-dimensional manifold is justified by the properties of the Laplace-Beltrami operator.

It is simple to illustrate Laplacian-based methods on sets sampled from an euclidean space. We start by assuming that a set of objects $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N \subset \mathbb{R}^D\}$ is sampled from a manifold $\mathcal{M} \subset \mathbb{R}^d$ and that $d \ll D$. The idea of these dimensionality reduction techniques is to construct a mapping

$$\mathbf{f} : \mathcal{M} \rightarrow \mathbb{R}^d$$

such that three hypotheses are met

1. Preservation of the distance relationship: the learned manifold should preserve the distance relation, two elements which were close in the original setting should stay close.
2. Uniform sampling of the elements: the manifold \mathcal{M} was uniformly sampled in order to produce X .
3. Convexity of the sampling: if two elements are in X , the elements between

those in \mathcal{M} have been sampled and those samples are in X .

Using the previous hypotheses, we formulate a mapping between two spaces as we did for the previous manifold learning techniques. However, instead of finding a discrete map for a sample of points, we start by obtaining a map between a continuous manifold and an euclidean space. This constitutes the theoretical grounds of Laplacian-Based methods. Finally, having this map for the continuous case we infer a mapping for the discrete case.

Finding the mapping for a manifold

In this section, we justify why the eigenvectors of the Laplace-Beltrami operator have desirable properties for the embedding. Following Belkin [2003], for the sake of simplicity we illustrate this by mapping a manifold embedded in \mathbb{R}^D to a line, up to some technical details, this procedure can be easily generalized to higher-dimensional mappings.

We start by stating that due to hypothesis 3, convexity, it is possible to bound the distance on the line with the manifold. For this, we use the previous approximation and we assume that the mapping $f(\cdot)$ exists and is twice differentiable

$$|f(\mathbf{x}) - f(\mathbf{z})| \leq \|\nabla f(\mathbf{x})\| \text{dist}_{\mathcal{M}}(\mathbf{x}, \mathbf{z}) + o(\text{dist}_{\mathcal{M}}(\mathbf{x}, \mathbf{z})). \quad (4.5)$$

Then, according to Belkin [2003], hypothesis 1 can be formulated as requiring that \mathcal{M} is isometrically embedded in \mathbb{R}^D . This leads to state that the geodesic distance in \mathcal{M} can be approximated up to a certain linear error by the euclidean distance in \mathbb{R}^D ,

$$\text{dist}_{\mathcal{M}}(\mathbf{x}, \mathbf{z}) = \|\mathbf{x} - \mathbf{z}\|_{\mathbb{R}^D} + o(\|\mathbf{x} - \mathbf{z}\|_{\mathbb{R}^D}), \quad \mathbf{x}, \mathbf{z} \in \mathcal{M}.$$

Merging the previous equation in equation 4.5 we obtain a bound on the mapping in terms of the original space where \mathcal{M} is embedded

$$|f(\mathbf{x}) - f(\mathbf{z})| \leq \|\nabla f(\mathbf{x})\| \|\mathbf{x} - \mathbf{z}\|_{\mathbb{R}^D} + o(\|\mathbf{x} - \mathbf{z}\|_{\mathbb{R}^D}).$$

According to the previous equation, the gradient of the mapping, $\nabla f(\cdot)$, is a measure of how apart $f(\cdot)$ maps nearby points of \mathcal{M} . Therefore, the mapping that best preserves the locality on average reduces the norm of the gradient,

$$f = \underset{\|f\|_{L^2(\mathcal{M})}=1}{\text{argmin}} \int_{\mathcal{M}} \|\nabla f\|^2$$

where the condition on $\|f\|$ is stated in order to remove the scale indetermination.

In order to find an explicit expression for $f(\cdot)$ through an eigenvalue problem, we state the previous equation in terms of the Laplace-Beltrami operator. We start by

using Stoke's theorem: if \mathbf{V} is a vector field

$$\int_{\mathcal{M}} \langle \mathbf{V}, \nabla f \rangle = - \int_{\mathcal{M}} \operatorname{div}(\mathbf{V})f,$$

this takes us to

$$\int_{\mathcal{M}} \|\nabla f\|^2 = \int_{\mathcal{M}} \langle \nabla f, \nabla f \rangle = \int_{\mathcal{M}} -\operatorname{div}(\nabla f)f,$$

where

$$\Delta(f) := -\operatorname{div}(\nabla f)$$

is called the *Laplace-Beltrami* operator over the manifold \mathcal{M} .

Hence, the mapping we are looking for is the function that minimizes the *Laplace-Beltrami* operator:

$$f = \operatorname{argmin}_{\|f\|_{L^2(\mathcal{M})}=1} \int_{\mathcal{M}} \Delta(f)f, \quad (4.6)$$

where $\Delta(f)$ is positive definite.

Equation 4.6, leads to two useful properties [Rosenberg: 1997]: if \mathcal{M} is compact and positive semi-definite, $\Delta(f)$ has a discrete set of eigenvalues $0 \leq \lambda_1 \leq \lambda_2 \leq \dots$ with its corresponding eigenfunctions $v_1(\cdot), v_2(\cdot), \dots$ where $v_i : \mathcal{M} \rightarrow \mathbb{R}$; and the function f minimizing $\int_{\mathcal{M}} \Delta(f)$ is an eigenfunction of Δ . Moreover, if \mathcal{M} is connected, $\lambda_1 = 0$ and the eigenfunction $v_1(\cdot)$ maps the entire manifold to one point. Finally, applying the same iterative reasoning we used in PCA to obtain an d -dimensional mapping, we reach our goal: an explicit expression for the d -dimensional mapping $\mathbf{f}(\cdot)$ of \mathcal{M} into the euclidean space of \mathbb{R}^d :

$$\mathbf{f}(\mathbf{x}) := (\lambda_2 v_2(\mathbf{x}), \dots, \lambda_{d+1} v_{d+1}(\mathbf{x})) = \mathcal{P}_d^{\text{LBM}} \in \mathbb{R}^d, \mathbf{x} \in \mathcal{M}.$$

In this mapping, its noticeable that we are using the smallest eigenvalues instead of the largest, as we do with PCA. This is due to the fact that the mapping $\mathbf{f}(\cdot)$, is the eigenfunction *minimizing* its product with respect to the Laplacian, as seen in equation 4.6, while in the PCA setting we look for the maximum solution of the eigenproblem.

We have just stated that given a manifold \mathcal{M} of dimension d embedded in D , it is possible to obtain a low dimensional mapping $\mathbf{f}(\cdot)$ by calculating the eigenvalues and eigenfunctions of its Laplace-Beltrami operator, Δ . The following step is to use this result in order to perform manifold learning on a set of points X .

Discrete Laplace-Beltrami operator

Getting back to the discrete scenario, we assume that we have a set of elements $X = \{x_1, \dots, x_N\}$ and a similarity function between them, $s(\cdot, \cdot)$, which, for theoretical correctness should be an inner product function. Providing that the hypotheses

stated at the beginning of this section are met, we can build a similarity matrix, or affinity matrix in the Laplace-based manifold learning literature,

$$\mathbf{S}_{ij} = s(\mathbf{x}_i, \mathbf{x}_j).$$

Then, taking \mathbf{S} as the connectivity matrix of a graph $G(X, E)$, the Laplacian of that graph is [Hein et al.: 2007]

$$\mathbf{L} = \mathbf{D} - \mathbf{S} \quad (4.7)$$

where \mathbf{D} , the degree matrix of the graph G , is defined as the diagonal matrix such that $\mathbf{D}_{ii} = \sum_j \mathbf{S}_{ij}$. Multiplying equation 4.7 by the vector $\mathbf{Y} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)]^T$, we note that the Laplace-Beltrami operator of $f(\cdot)$ for an element of X is approximated in this case as:

$$(\mathbf{L}f)(\mathbf{x}_i) = \sum_{j=1}^N s(\mathbf{x}_i, \mathbf{x}_j) f(\mathbf{x}_i) - \frac{1}{N} \sum_{j=1}^N s(\mathbf{x}_i, \mathbf{x}_j) f(\mathbf{x}_j)$$

The previous approximation to the Laplacian is known as the *unnormalized* Graph Laplacian. In usual literature there are at least two more ways of defining this approximation: the normalized, \mathbf{L}_n , and random walk, \mathbf{L}_r , Graph Laplacians [Hein et al.: 2007]:

$$\begin{aligned} \mathbf{L}_n = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{S} \mathbf{D}^{-\frac{1}{2}} & \quad (\mathbf{L}_n f)(\mathbf{x}_i) = f(\mathbf{x}_i) - \frac{1}{N} \sum_{j=1}^N \frac{s(\mathbf{x}_i, \mathbf{x}_j) f(\mathbf{x}_j)}{\sqrt{\sum_{k=1}^N s(\mathbf{x}_i, \mathbf{x}_k)} \sqrt{\sum_{k=1}^N s(\mathbf{x}_j, \mathbf{x}_k)}} \\ \mathbf{L}_r = \mathbf{I} - \mathbf{D}^{-1} \mathbf{S} & \quad (\mathbf{L}_r f)(\mathbf{x}_i) = f(\mathbf{x}_i) - \frac{1}{N} \frac{\sum_{j=1}^N s(\mathbf{x}_i, \mathbf{x}_j) f(\mathbf{x}_j)}{\sum_{k=1}^N s(\mathbf{x}_i, \mathbf{x}_k)} \end{aligned}$$

The difference between these three definitions is their convergence towards the Laplace-Beltrami operator when the number of sampled elements from the manifold tends to infinity.

Normalization and convergence

Hein et al. [2007] analysed the convergence of the most usual graph Laplacian operators, a work that had been started by Lafon and Lee [2006]. The main idea of this section is to study how the different versions of the graph Laplacian converge towards the Laplace-Beltrami operator as the sampling of the manifold \mathcal{M} which produces the set of elements to be embedded X , increases.

To study the convergence of the Graph Laplacians, we provide \mathcal{M} with a non-uniform probability measure which represents the density of the manifold sampling, $p(\cdot)$. Using this measure, Hein et al. [2007], rewrite the Laplace-Beltrami operator as

$$\Delta_s = \frac{1}{p^s} \operatorname{div}(p^s \nabla) \quad (4.8)$$

they consider that the points X are sampled from \mathcal{M} under an unknown density $p(\cdot)$ and that $N \leq D$ meaning we have at least as many samples as the number of dimensions of \mathcal{M} . Having a Laplace-Beltrami operator depending on a power of the sampling function, we construct the similarity function such that it is normalized by the empirical density of the sampling:

$$\tilde{s}_\beta(\mathbf{x}_i, \mathbf{x}_j) = \frac{s(\mathbf{x}_i, \mathbf{x}_j)}{(q(\mathbf{x}_i)q(\mathbf{x}_j))^\beta}$$

where the normalizing function $q(\mathbf{x}) = \sum_{\mathbf{y} \in X} s(\mathbf{x}, \mathbf{y})$ is the Nadaraya-Watson estimate of density $p(\cdot)$ at \mathbf{x} and β regulates the decay the similarity measure [Hein et al.: 2007]. The two scalars β and s are powers of the density function, and can be linked as $\beta = 1 - s/2$. Then, density-independent versions of the discrete Laplacian, $\tilde{\mathbf{L}}$, $\tilde{\mathbf{L}}_n$, $\tilde{\mathbf{L}}_r$ are built by simply changing \mathbf{S} by $\tilde{\mathbf{S}}_{ij} = \tilde{s}_\beta(\mathbf{x}_i, \mathbf{x}_j)$. Hein et al. [2007] calculated the almost sure convergence of these discrete Laplacians as the quantity of elements of X , N , tends to infinity:

$$\tilde{\mathbf{L}}(x) \rightarrow -p(x)^{1-2\beta}(\Delta_{2(1-\beta)}\mathbf{f})(x) \quad (4.9)$$

$$\tilde{\mathbf{L}}_n(x) \rightarrow -p(x)^{\frac{1}{2}-\beta} \left(\Delta_{2(1-\beta)} \frac{\mathbf{f}}{p(\mathbf{x})^{\frac{1}{2}-\beta}} \right) (x) \quad (4.10)$$

$$\tilde{\mathbf{L}}_r(x) \rightarrow -(\Delta_{2(1-\beta)}\mathbf{f})(x) \quad (4.11)$$

The first observation is that the graph Laplacian converges to the Laplace-Beltrami operator only when $p(\cdot)$ is constant, or equivalently, when the manifold was sampled uniformly as stated by the hypotheses at the beginning of this section. In the generic case of manifold learning, when this can not be assured, there are big differences among the discretised Laplacian operators. The only one of these which converges towards a weighted Laplace-Beltrami operator in this case is the random walk version, $\tilde{\mathbf{L}}_r$. However, the limit of the unnormalized Laplacian, $\tilde{\mathbf{L}}$ is useful in manifold learning applications: the extra $p(\cdot)^{1-2\beta}$ actually means that if $\beta < \frac{1}{2}$, the distance propagates faster in regions when the density is high and will render them closer in the embedding. In their work, Hein et al. [2007] could not provide an interpretation or scenario where the normalized Laplacian was useful. However von Luxburg et al. [2005] proved that in the particular case of clustering, the normalized Laplacian performs better when the main goal of the embedding is a posterior clustering application, most probably due to its equivalence to the Laplacian operator proposed by Shi and Malik [2000] to perform Normalized-cuts clustering.

Finally, analysing the convergence of the three versions of the discrete Laplacian operator it is possible draw conclusions about their uses cases. First, when sampling density should be reflected by the embedding, the unnormalized Laplacian is applicable, however if no conditions can be assured from the original density and the user wants the embedding to be independent of this, the random walk version should be

used.

Laplacian Eigenmaps algorithm

Based on the previous theory, [Belkin and Niyogi \[2003\]](#) proposed to use the graph Laplacian in order to perform manifold learning. Given a set of elements X and a dissimilarity function between them, $d(\cdot, \cdot)$, they proposed to turn the dissimilarity into a similarity measure using a kernel function:

$$s_\sigma(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{d^2(\mathbf{x}_i, \mathbf{x}_j)}{\sigma^2}\right)$$

where σ is the kernel bandwidth. Having this done, they compute the similarity matrix

$$\mathbf{S}_{ij} = s_\sigma(\mathbf{x}_i, \mathbf{x}_j)$$

and use the unnormalized Laplacian, equation 4.7, to perform the embedding. The technique presented in this section, called Laplacian Eigenmap manifold Learning (LEM), is described in algorithm 4.1.

In their work, [Belkin and Niyogi \[2003\]](#) show that the graph-theoretic spectral clustering algorithms, described in subsection 3.2.2, are actually performing Laplacian-based embedding. These algorithms perform the embedding either with the unnormalized or normalized Laplacian and then applying a clustering algorithm on the embedded data. Finally, the main advantage with respect to Isomap is the fact that they preserve local instead of global topology. The advantages of this are twofold: these embedding techniques tolerate better the curvature on the original manifold M ; they produce sparser matrices reducing the computational cost of the eigendecomposition.

Diffusion Maps algorithm

[Lafon \[Lafon: 2004, Lafon and Lee: 2006, Coifman et al.: 2005\]](#) proposed the Diffusion Maps (DM) algorithm in order to overcome some shortcomings of the previous techniques, in particular LEM. The main idea behind Diffusion Maps is to provide an embedding which is resilient to differences on the sampling density of X from \mathcal{M} . As we have seen in subsection 4.2.3, this can be done in one of three ways: using the random walk graph Laplacian, \mathbf{L}_r ; or setting β to $\frac{1}{2}$ and using the normalized \mathbf{L}_n or unnormalized Laplacian, \mathbf{L} .

In his work, Lafon explored the use of these three settings and proposed two techniques. The first technique, shown in algorithm 4.2 uses the normalized Laplacian [[Coifman et al.: 2005](#)] and the second one, the random walk Laplacian [[Lafon and Lee: 2006](#)], described in algorithm 4.3. As we have seen in subsection 4.2.3, the normalized Laplacian is related to the NCuts segmentation algorithm, chapter 3: having as a goal the balance of the density of connections in the graph in order to perform

Algorithm 4.1 Laplacian Eigenmaps manifold learning [Belkin and Niyogi: 2003]

Inputs A set of elements to embed $X = \{x_1, \dots, x_N\}$, a dissimilarity function $d(\cdot, \cdot)$ and a bandwidth parameter σ

Outputs A d dimensional map for every element of X .

1: Build the similarity matrix

$$\mathbf{S}_{ij} = \exp\left(\frac{d^2(\mathbf{x}_i, \mathbf{x}_j)}{\sigma^2}\right)$$

2: Optionally, set so zero elements of \mathbf{S} (prune the graph) according to the ϵ or k neighbours criteria in subsection 4.2.2.

3: Compute the discrete unnormalized Laplacian

$$\mathbf{L} = \mathbf{D} - \mathbf{S}$$

where $\mathbf{D} = \text{diag}(\sum_i \mathbf{S}_{1i}, \dots, \sum_i \mathbf{S}_{Ni})$

4: Perform the eigendecomposition of \mathbf{L} in the eigenvalues $0 = \lambda_1 \leq \dots \leq \lambda_M$ and eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_N$

5: **return** the d -dimensional mapping for \mathbf{x}_i :

$$\mathcal{P}_d^{\text{LEM}}(\mathbf{x}_i) = ((\mathbf{v}_2)_i, \dots, (\mathbf{v}_{d+1})_i)$$

better for clustering algorithms. The version using the random walk Laplacian, \mathbf{L}_r has the advantage which is sustained in the random-walk interpretation of \mathbf{L}_r ,

$$\mathbf{L}_r = \mathbf{I} - \mathbf{D}^{-1}\mathbf{S}.$$

In this case, the matrix $\mathbf{P} = \mathbf{D}^{-1}\mathbf{S}$ is called its diagonal dominant and its rows add up to 1. This leads to the interpretation of the matrix entry \mathbf{P}_{ij} as the probability that a random walker starting from \mathbf{x}_i ends up in \mathbf{x}_j after a time step. Moreover, if we note this probability as $\mathbb{P}\{x_1 = \mathbf{x}_j | x_0 = \mathbf{x}_i\} = \mathbf{P}_{ij}$, then we can measure the probability after several time-steps as [Lovász: 1993]

$$\mathbb{P}\{x_t = \mathbf{x}_j | x_0 = \mathbf{x}_i\} = \mathbf{P}_{ij}^t.$$

From the fact that

$$\mathbf{P} = \mathbf{I} - \mathbf{L}_r$$

it is simple to deduce that the i -th eigenvalues of \mathbf{P} is $1 - \lambda_i$, where λ_i is the i -th eigenvalue of \mathbf{L}_r . Hence, the distance between the mapping

$$\mathbf{f}_t(\mathbf{x}_i) = \left(\lambda_2^t \frac{(\mathbf{v}_2)_i}{(\mathbf{v}_1)_i}, \dots, \lambda_{d+1}^t \frac{(\mathbf{v}_{d+1})_i}{(\mathbf{v}_1)_i} \right)$$

of two elements

$$d_t^{\text{DM}}(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{f}_t(\mathbf{x}_i) - \mathbf{f}_t(\mathbf{x}_j)\|_2$$

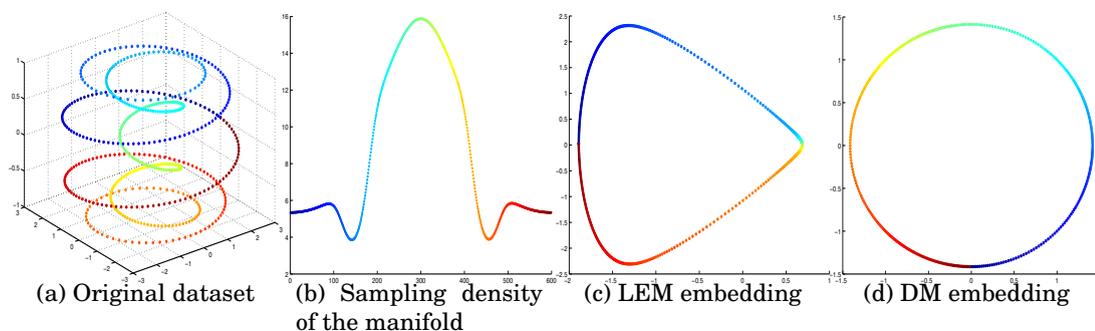


Figure 4.6: Laplacian-based manifold learning techniques and sampling-induced effects. From right to left: the original set to be embedded in a two-dimensional manifold; the empirical sampling density of the manifold; the embedding obtained by Laplacian-eigenmaps, it is noticeable how the curvature of the embedding changes in sections of the manifold where the sampling has a local maxima or minima; and diffusion maps-based embedding where the differences in the sampling had no effect. Adapted from [Lafon: 2004]

or diffusion distance, can be interpreted as a measure of the connectivity of the elements x_i and x_j after a time t . Moreover, the parameter t can be changed in order to change the scale of the embedding after the embedding was performed.

The goal of this contribution is to develop manifold learning techniques for a posterior clustering of the white matter bundles. Due to this, the version of DM based on the normalized Laplacian is the one we will chose as a part of our clustering algorithm. The results of an embedding using this technique compared with LEM is shown in figure 4.6, where we observe that DM is resilient to sampling differences.

4.2.4 Other techniques

Other techniques have been proposed for manifold learning. Roweis and Saul [2000] proposed the Locally Linear Embedding, which stands in middle ground between Isomap and the Laplacian Eigenmaps as it performs explicit calculation of the tangent planes of the manifold instead of the geodesics. This technique was later extended by Donoho and Grimes [2003] as Hessian Eigenmaps. The main idea of this technique is to relax the convexity hypothesis by of performing a second order approximation to the manifold instead of a first order one. Several methods have been published in the last ten years but without the impact of the ones presented in this chapter. We refer the reader to *computer vision* or *machine learning* literatures in which new techniques appear frequently.

4.2.5 Estimating the Dimension of the Manifold

The dimension of a manifold is also known as its intrinsic dimension. The chosen dimension should be the minimal number of parameters necessary to represent the variability of the dataset up to the point to which the subsequent analysis of the data needs. For instance, for a clustering technique, a number of dimensions allowing a

Algorithm 4.2 Diffusion Maps using the normalized Laplacian [Coifman et al.: 2005]

Inputs A set of elements to embed $X = \{x_1, \dots, x_N\}$, a dissimilarity function $d(\cdot, \cdot)$ and a bandwidth parameter σ and a diffusion distance t

Outputs A d dimensional map for every element of X .

- 1: Build the similarity matrix

$$\mathbf{S}_{ij} = \exp\left(\frac{d^2(\mathbf{x}_i, \mathbf{x}_j)}{\sigma^2}\right)$$

- 2: Normalize \mathbf{S} according to its density

$$\mathbf{P} = [\mathbf{D}(\mathbf{S})]^{-1} \mathbf{S} [\mathbf{D}(\mathbf{S})]^{-1}$$

where $\mathbf{D}(\mathbf{S}) = \text{diag}(\sum_i \mathbf{S}_{1i}, \dots, \sum_i \mathbf{S}_{Ni})$

- 3: Build the normalized Laplacian of the graph induced by \mathbf{P}

$$\mathbf{L}_n = \mathbf{I} - [\mathbf{D}(\mathbf{P})]^{-\frac{1}{2}} \mathbf{P} [\mathbf{D}(\mathbf{P})]^{-\frac{1}{2}}$$

- 4: Perform the eigendecomposition of \mathbf{L}_n in the eigenvalues $0 = \lambda_1 \leq \dots \leq \lambda_M$ and eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_N$

- 5: **return** the d -dimensional mapping for \mathbf{x}_i :

$$\mathcal{P}_d^{\text{DM}}(\mathbf{x}_i) = \left(\lambda_2 \frac{(\mathbf{v}_2)_i}{(\mathbf{v}_1)_i}, \dots, \lambda_{d+1} \frac{(\mathbf{v}_{d+1})_i}{(\mathbf{v}_1)_i} \right)$$

clustering algorithm to find a partition of the data adjusted to the users necessities is enough.

A first approach to estimate the dimensionality is observing the eigenspectrum, the spectrum of the eigenvalues output by the eigendecomposition at the final step of all the algorithms. Take for instance PCA, in this case, the magnitude of each eigenvalue represents the amount of useful information which corresponds to its the dimension associated to its eigenvector. Then, in ideal cases, the number of dimensions corresponds to the number of non-zero eigenvalues.

In less trivial scenarios, there are several heuristics that can be used. The most frequently used is the gap heuristic [Belkin and Niyogi: 2003, Coifman et al.: 2005]: Consider the eigenvalues sorted in decreasing order of magnitude, $\lambda_1 \geq \lambda_2 \geq \dots$, if the i -th eigenvalue shows an abrupt change of slope, then the number of dimensions to use is i . This heuristic formalized by calculating the dimensionality d^* as

$$d^* = \arg \min_d \left| \frac{\sum_{i=1}^d \lambda_i}{\sum_{i>d} \lambda_i} - x \right|$$

where x is a cut-off value which must be arbitrarily chosen and quantifies how abrupt the change of slope must be.

More formal approaches have been developed recently in order to calculate the

Algorithm 4.3 Diffusion Maps using the random walk Laplacian [[Lafon and Lee: 2006](#)]

Inputs A set of elements to embed $X = \{x_1, \dots, x_N\}$, a dissimilarity function $d(\cdot, \cdot)$ and a bandwidth parameter σ and a diffusion distance t

Outputs A d dimensional map for every element of X .

- 1: Build the similarity matrix

$$\mathbf{S}_{ij} = \exp\left(\frac{d^2(\mathbf{x}_i, \mathbf{x}_j)}{\sigma^2}\right)$$

- 2: Compute the discrete random walk Laplacian

$$\mathbf{L}_r = \mathbf{I} - \mathbf{D}^{-1}\mathbf{S}$$

where $\mathbf{D} = \text{diag}(\sum_i \mathbf{S}_{1i}, \dots, \sum_i \mathbf{S}_{Ni})$

- 3: Perform the eigendecomposition of \mathbf{L}_r in the eigenvalues $0 = \lambda_1 \leq \dots \leq \lambda_M$ and eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_N$
- 4: **return** the d -dimensional mapping for \mathbf{x}_i :

$$\mathcal{P}_{d,t}^{\text{DMRW}}(\mathbf{x}_i) = \left(\lambda_2^t \frac{(\mathbf{v}_2)_i}{(\mathbf{v}_1)_i}, \dots, \lambda_{d+1}^t \frac{(\mathbf{v}_{d+1})_i}{(\mathbf{v}_1)_i} \right)$$

intrinsic dimensionality of data, for instance [Levina and Bickel \[2004\]](#) and [Hein and Audibert \[2005\]](#) and the references therein.

Although existing techniques could be used to determine the dimension of our manifold, we mostly set it as a parameter. Moreover, in the particular application of clustering it has been argued that the number of clusters to detect in the data is equal to the number of dimensions which parametrize the manifold where data lies [[Zelnik-Manor and Perona: 2004](#), [Lee and Wasserman: 2009](#)]¹.

4.3 SUMMARY

In section subsection [3.1.1](#) we had seen that the representation of the proximity is of great importance for clustering tasks. At the beginning of this chapter, we analysed the case where an evident proximity measure, or explicit, does not represent accurately the elements to analyse through clustering leading to incorrect results. This motivated the introduction and analysis of manifold learning techniques, designed specifically in order to learn an implicit parametrization of the dataset to cluster hence providing correct clustering methods. In the following chapter we apply manifold learning and clustering techniques in order to identify white matter structures from dMRI images.

1. Note the lacking n in the second author's last name, it is not me.

WHITE MATTER STRUCTURE IDENTIFICATION

Having set the basis for manifold learning and clustering, we proceed to implement a Diffusion-Maps based algorithm to perform automatic identification of white matter fibre bundles. This algorithm has several advantages over its predecessors, We achieve the segmentation through spectral clustering. These techniques have been successfully applied to image processing since [Shi and Malik: 2000], providing a method for image processing which involves manifold learning and it is initialization free. In order to achieve this, a metric between Q-Ball ODFs is needed. Which metric should be used on Q-Ball image analysis and clustering is still an open metric.

In this chapter, we describe a clustering algorithm that infers an embedding and a metric to compare ODF images. We derive a similarity measure incorporating the Euclidean distance and the spatial location distance between ODFs. Then, we show that the Q-ball ODF clustering using Diffusion Maps can reproduce the DT clustering on simple synthetic images without crossings. On more complex data with crossings, we show that our method succeeds to separate fibre bundles and crossing regions on synthetic data, where the DT-based methods generate artifacts and exhibit wrong number of clusters. Finally, we successfully segment the fibre bundles in a real human brain dataset in different regions with fibres crossing.

5.1 SPECTRAL EMBEDDING AND CLUSTERING

As we have seen in subsection 4.1.1, non-linear manifold-learning techniques, more precisely Diffusion Maps take a similarity matrix as input. Hence, we proceed to describe the construction of the similarity matrix for white matter bundles segmentation.

5.1.1 Distance Functions Between Elements to Cluster

In order to implement the Diffusion Maps spectral clustering method a distance function for each data type is chosen. This distance function is used to calculate the similarity matrix as expressed by algorithm 4.2. In the DT case we use the tensor

distance shown in equation 2.8. In the ODF case we use the distance shown in equation 2.16.

5.1.2 Similarity Matrix Construction and Diffusion Maps Embedding

Each element of a DT or a ODF image can be represented as having two components. In the DT case (\mathbf{x}, \mathbf{D}) , where $\mathbf{x} \in \mathbb{R}^3$ is the position in the image and $\mathbf{D} \in \mathbb{R}^{3 \times 3}$ symmetric, positive definite. In the Q-Ball case, (\mathbf{x}, Ψ) , where $\Psi \in \mathbb{R}^L$ are the spherical harmonic coefficients. In both cases there is two kinds of information to be taking in account, the positional one and the representing the diffusion at each position. Both information kinds need to be taking in account differently. In the image segmentation case [Shi and Malik: 2000] propose a solution by taking, as a distance, a weighted sum of two separate distances, in this case the positional and the diffusivity one, but this adds a parameter to the algorithm and the way of mixing both distances is not intuitive. In this work, we chose a random walk based technique. This technique is more cohesive with the spectral embedding approach, which has been related with random walks by [Lafon and Lee: 2006]. Following [Tishby and Slonim: 2000, Ziyang et al.: 2006], we use the similarity between the elements in two neighboring voxels as a way to calculate the probability of a random walker going from one voxel to the other. Then a matrix relating only neighboring voxels through its walking probability can be built as,

$$\{\mathbf{S}_1\}_{ij} = \begin{cases} s(\mathcal{D}_i, \mathcal{D}_j) & \text{if } \|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq 1 \\ 0 & \text{any other case} \end{cases}$$

where $\mathcal{D}_i, \mathcal{D}_j$ is the diffusivity part- \mathbf{D} or Ψ -of the two elements to be clustered, $s(\mathcal{D}_i, \mathcal{D}_j)$ is defined as in algorithm 4.2, and \mathbf{x}_i are spatial coordinates of element \mathcal{D}_i in the image. In order to turn the affinities into probabilities, the matrix \mathbf{S}_1 must be normalized,

$$\mathbf{P} = \frac{1}{\max_l D(\mathbf{S}_1)_{ll}} \begin{cases} \max_l D(\mathbf{S}_1)_{ll} - D(\mathbf{S}_1)_{ii} & \text{if } i = j \\ \{\mathbf{S}_1\}_{ij} & \text{any other case} \end{cases},$$

where diagonal adjustment forces the inherent random walk to a uniform steady state, hence every part of the Markov field will be explored at the same speed. Then, $(\mathbf{P}_{ij})^t, t = \{1, 2, \dots\}$ represents the probability of arriving to the voxel \mathbf{x}_j , starting from \mathbf{x}_i through walking t steps.

In order to illustrate this, we show a simple example with a scalar image. The similarity is calculated as $s(\mathcal{D}_i, \mathcal{D}_j)$, where the distance is the absolute difference between the intensity on the image. The scalar image is figure 5.1a, the probabilities of reaching other voxels in the image with one step starting from a voxel in the edge

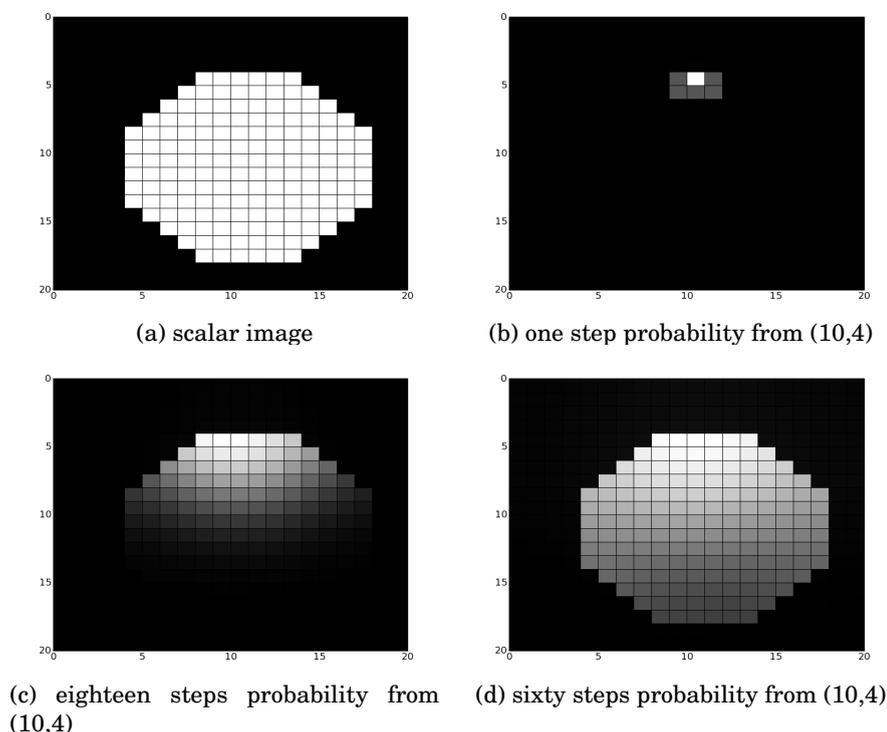


Figure 5.1: Markovian relaxation to calculate connectivity in an image. From a scalar image figure 5.1a, the probabilities of reaching other voxels in the image with one step starting from a voxel in the edge of the circle are shown in figure 5.1b. Then the probabilities of reaching other voxels in the image with 18 steps starting from a voxel in the edge of the circle are shown in figure 5.1c and with 60 steps figure 5.1d

of the circle are shown in figure 5.1b. As each voxel is an element of X , this is equivalent as calculating P for the image and taking P_{ij} as the probability of starting in i and reaching j , thus figure 5.1b is obtained by mapping the matrix row P_i to a bi-dimensional image. Finally, the probabilities of reaching every voxels in the image by taking 18 and 60 steps starting from a voxel in the edge of the circle are shown in figure 5.1c and figure 5.1d, where the values are taken from the matrix P^{18} and P^{60} respectively.

Then, the matrix P^t is taken as the input for the diffusion maps embedding process described in algorithm 4.2. We chose the version of diffusion maps based on the normalized Laplacian due to the studies on its convergence and on the balancing of cluster size that we described in chapters 3 and 4. The quanta of steps t is chosen to be the smallest integer where P^t does not have any element with a value of 0, giving in this way the weakest connected random walk over the whole image.

As an output of the Diffusion Maps algorithm, we obtain a set of elements $Y = \{y_1, \dots, y_N\} \subset \mathbb{R}^D$, each element represents a voxel and the euclidean distance between these elements represents their similarity, taking in account their spatial and diffusivity information and their dimensionality is left to be chosen in during the clustering stage, according to subsection 4.2.5.

5.1.3 Clustering

Once the embedding has been performed, several techniques have been proposed for the clustering step. [Shi and Malik \[2000\]](#), [Belkin and Niyogi \[2003\]](#), [Lafon and Lee \[2006\]](#)

The first step in this process is to determine the number of clusters, this can be done in two ways. The first, as done by [\[Nadler et al.: 2006\]](#), is choosing the number of clusters according to the “elbow” present in the eigenvalue plot. For instance, if the slope of the eigenvalue plot changes noticeably at eigenvector λ_i , the number of clusters should be $i + 1$. The second way is re-ordering the affinity matrix rows and columns following the second eigenvector, as proved by [\[Fiedler: 1975\]](#), which shows the block structure of the matrix as squared blocks along the matrix diagonal. Then, the number of clusters is the number of blocks. The recommended number of dimensions for the embedding is the same as the number of clusters. Finally, the clustering is performed by running a k-means clustering algorithm on y , an approach previously taken by [Belkin and Niyogi \[2003\]](#) and [Lafon and Lee \[2006\]](#).

5.2 MATERIALS

5.2.1 Synthetic Data

We generate synthetic HARDI data using the multi-tensor model which is simple and leads to an analytical expression of the ODF [\[Descoteaux et al.: 2007a, Tuch: 2004\]](#). For a given b -factor and noise level, we generate the diffusion-weighted signal

$$S(\mathbf{u}_i) = \sum_{k=1}^n \frac{1}{n} \exp(-b\mathbf{u}_i^T \mathbf{D}_k(\theta) \mathbf{u}_i) + noise \quad (5.1)$$

where \mathbf{u}_i is the i^{th} gradient direction on the sphere, n is the number of fibres and $1/n$ is the volume fraction of the of each fibre. In practice, we use $N = 81$ from a 3^{rd} order tessellation of the icosahedron, $b = 3000 \text{ s/mm}^2$ and $n = 1$ or 2 . $\mathbf{D}_k(\theta)$ is the diffusion tensor with standard eigenvalues $[3, 3, 1.7] \times 10^{-2} \text{ mm}^2/\text{s}$ oriented in direction θ , which agree with reported physiological values [\[Pierpaoli et al.: 1996\]](#). Finally, we add complex Gaussian noise with standard deviation of $1/35$, producing a signal with signal to noise ratio (SNR) of 35.

We generate three synthetic data example, two simple examples, one with a ring of sinusoidal shaped fibres, one with fibres with different sizes and scales and the other with complex crossing areas simulating the ‘U’-fibres (cortico-cortical fibres) that can occur in the brain. These synthetic datasets help understand the behavior of the different spectral clustering methods when confronted with simple and complex fibre geometries.

5.2.2 Human Brain Data

Diffusion weighted data and high-resolution T1-weighted images were acquired on a whole-body 3 Tesla Magnetom Trio scanner (Siemens, Erlangen) equipped with an 8-channel head array coil [Anwander et al.: 2007]. The spin-echo echo-planar-imaging sequence, TE = 100 ms, TR = 12 s, 128 x 128 image matrix, FOV = 220 x 220 mm², consists of 60 diffusion encoding gradients [Jones et al.: 1999] with a b-value of 1000 s/mm². Seven images without any diffusion weightings are placed at the beginning of the sequence and after each block of 10 diffusion weighted images as anatomical reference for offline motion correction. The measurement of 72 slices with 1.7mm thickness (no gap) covered the whole brain. Random variations in the data were reduced by averaging 3 acquisitions, resulting in an acquisition time of about 45 minutes. No cardiac gating was employed to limit the acquisition time. The issue of cardiac gating is discussed by Jones et al. [2002]. Additionally, fat saturation was employed and we used 6/8 partial Fourier imaging, a Hanning window filtering and parallel acquisition (generalized auto-calibrating partially parallel acquisitions, reduction factor = 2) in the axial plane.

The brain is peeled from the T1-anatomy, which was aligned with the Talairach stereotactical coordinate system [Talairach and Tournoux: 1988]. The 21 images without diffusion weightings distributed within the whole sequence were used to estimate motion correction parameters using rigid-body transformations [Jenkinson et al.: 2002], implemented in [FSL: 2006]. The motion correction for the 180 diffusion-weighted images was combined with a global registration to the T1 anatomy computed with the same method. The gradient direction for each volume was corrected using the rotation parameters. The registered images were interpolated to the new reference frame with an isotropic voxel resolution of 1.72 mm and the 3 corresponding acquisitions and gradient directions were averaged.

5.3 EXPERIMENTS AND RESULTS

5.3.1 Synthetic data experiments

Diffusion Maps vs. N-Cuts

The first experiment shows the difference in performance between the Diffusion Maps and N-Cuts approach. The N-Cut algorithm does not perform the sampling-based normalization described in subsection 4.2.3 and is thus sensitive to sampling frequency differences within the clusters. In order to show this sampling hypothesis problem, we used a ring fibre bundle with different sampling frequencies. Within the ring, the fibres have a sinusoidal shape and the frequency of the modulating sine function is 4 times bigger in the lower half of the ring. More formally, the fibres fol-

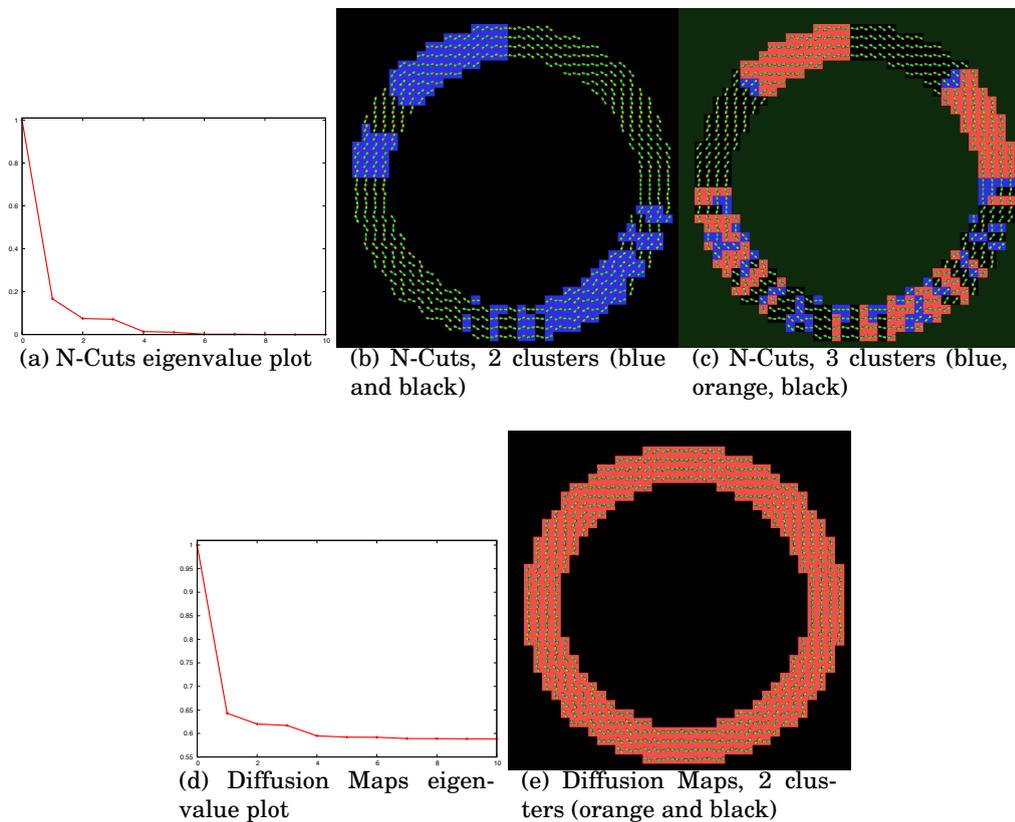


Figure 5.2: N-Cuts generates over-clustering due to sampling frequency variation in ODF images. In both eigenvalue plots figure 5.2a, figure 5.2d, the slope between the line joining λ_0 and λ_1 and the line joining λ_1 and λ_2 changes drastically, expressing an elbow in λ_1 , which indicates two clusters. The clustering results with 2, 5.2b, and 3, 5.2c, clusters are shown. Diffusion Maps correctly finds two clusters, the object and the background, 5.2e. In the labeling, the ODFs are overlaid on the labels.

low the angular function $o(\theta) = \theta + \frac{1}{8}\pi \sin(\mu \cdot \theta), 0 \leq \theta < 2\pi$, where $\mu = 8$ for the upper half of the ring and $\mu = 32$ for the lower half. Two clusters are expected, the ring and the background. The results of both clustering techniques are shown in figure 5.2, where the background has been masked out. Figure 5.2a shows the plot of the first 10 eigenvalues for the N-Cuts method, shown in figure 5.2b and figure 5.2c. The slope between the line joining λ_0 and λ_1 and the line joining λ_1 and λ_2 changes drastically. This elbow at λ_1 indicates that there are 2 clusters. Figure 5.2d shows the plot of the first 10 eigenvalues for the Diffusion Maps method whose clustering results are shown in figure 5.2e. The N-Cuts exhibits frequency-dependent clustering artifacts while the Diffusion Maps method clearly shows two clusters. In the Diffusion Maps, the clustering has correctly segmented the background and the ring.

ODF vs. DT images

In figure 5.3, a single fibre scenario with no fibre crossing is shown. The DT-based and ODF-based image clustering produce the same results. Hence, ODF clus-

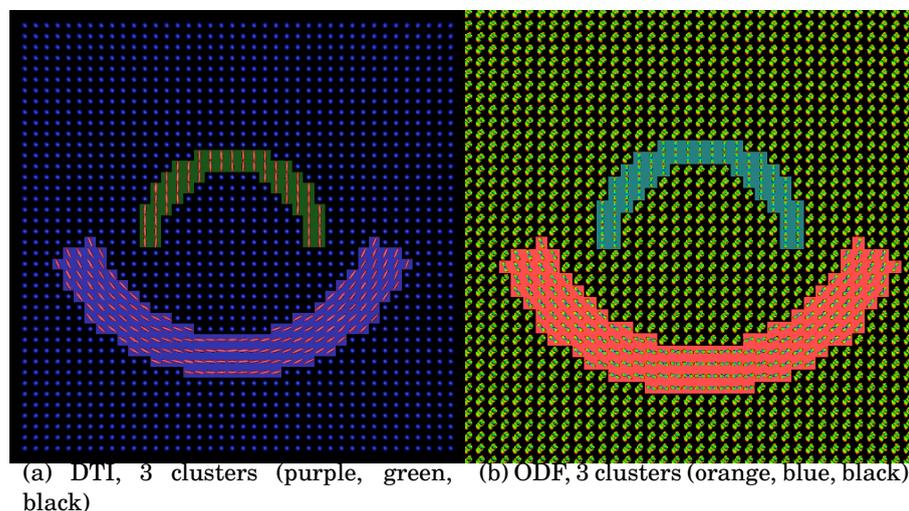


Figure 5.3: Synthetic image without fibre crossings. The results for the DT and ODF images are equivalent. The colors behind the DTs and ODFs indicate the clusters.

tering reproduces DT-based results on a simple fibre population example.

Finally, figure 5.4 shows a fibre crossing scenario with two overlapping fibre bundles that have different geometries. Segmentation was performed over the DT and the ODF image shown in figure 5.5. Note that the cluster number is correctly estimated only in the ODF image. Moreover, the ODF N-Cuts segmentation exhibits artifacts not present in the ODF Diffusion Maps segmentation. The ODF Diffusion Maps effectively identify the two different fibre bundles as well as the fibre crossing areas.

5.3.2 Real Data

The real data experiment presented in this section shows the segmentation and labeling of a cropped axial and coronal slice. The cropped slices were chosen such that they contain regions of known fibre crossings where the DT model is normally limited. The ROIs show intersection of several fibre bundles. Hence, our segmentation algorithm is confronted with elements that have different orientation and diffusion characteristics.

In order to show that ODF data segments the white matter fibre bundles better than the DT data in real cases, we analyse the evolution of the similarity matrix as the scale space parameter changes. The region where we extract the elements from is a cropped axial slice shown in figure 5.6. Similarity matrices were computed with varying scale space parameter between $\frac{1}{5}$, $\frac{1}{10}$, $\frac{1}{20}$ and $\frac{1}{40}$ of the quantity of elements ($|X|$) to cluster respectively. In order to show the block structure of these matrices, we reordered them using the second (Fiedler) largest eigenvector as explained in subsection 3.2.2 [Fiedler: 1975]. In figure 5.7 we can note that as the scale diminishes, the DT data shows a high correlation between all the elements of the slice.

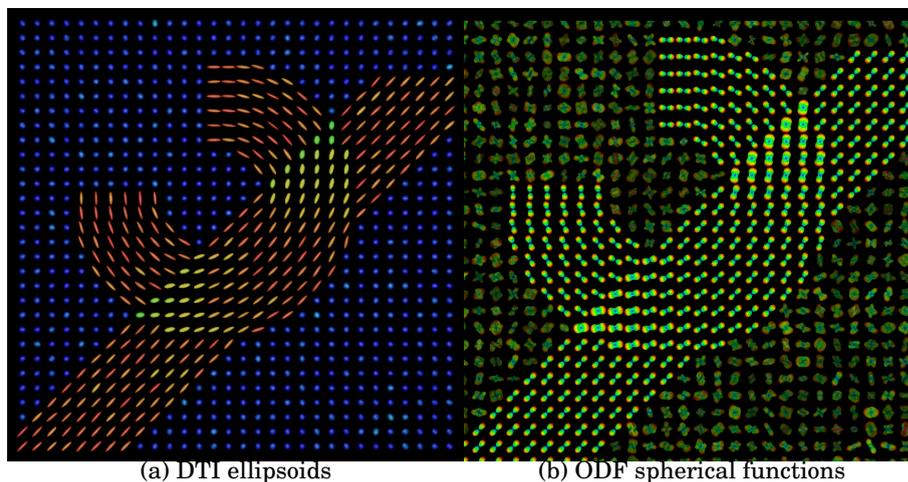


Figure 5.4: Synthetic DT and ODF images. The expected number is four, one for each fibre, one for the crossing between the two fibres and one for the background

This makes clustering very difficult because the blocks are small and highly correlated. On the other hand, the ODF data shows a very clear block structure across all scales. This structure accounts for a high correlation of the elements within each block and a low inter-block correlation, which indicates that the output of the manifold learning in this case is a much better input to the clustering algorithm than the DT data.

Then we apply the algorithm to an axial slice. Figure 5.6 shows the location of this slice in cropped axial (figure 5.6a) and coronal slices (figure 5.6b). We can observe in the segmented and labeled axial slice, figure 5.8, that the segmentation also succeeds in identifying some of the main white matter structures: Corpus Callosum (CC), Anterior Corona Radiata (ACR), Forceps Major (fmajor) and Forceps Minor (fminor).

In figure 5.9, the location of the cropped coronal slice is shown in the axial , figure 5.9a, and coronal slices, figure 5.9b. In the segmented coronal slice, figure 5.9c, we observe that the segmentation allows to identify and label main white matter structures: Corpus Callosum (CC), Cingulum (CG), Corona Radiata (CR), Superior Longitudinal Fasciculus (SLF). Note that the segmentation is resilient to crossing areas such as seen at the interface between CR and CC.

5.4 DISCUSSION

Results showed that our diffusion maps-based clustering algorithm for QBall imaging is a fit tool to segment white matter bundles. Firstly, we performed comparisons against techniques used in previous literature using synthetic images. We showed that the lack of uniform sampling from the manifold containing the ODFs can generate artefacts in NCuts-based segmentation while techniques based on diffusion maps are resilient to this sampling characteristics. Then, we showed that, in simple ar-

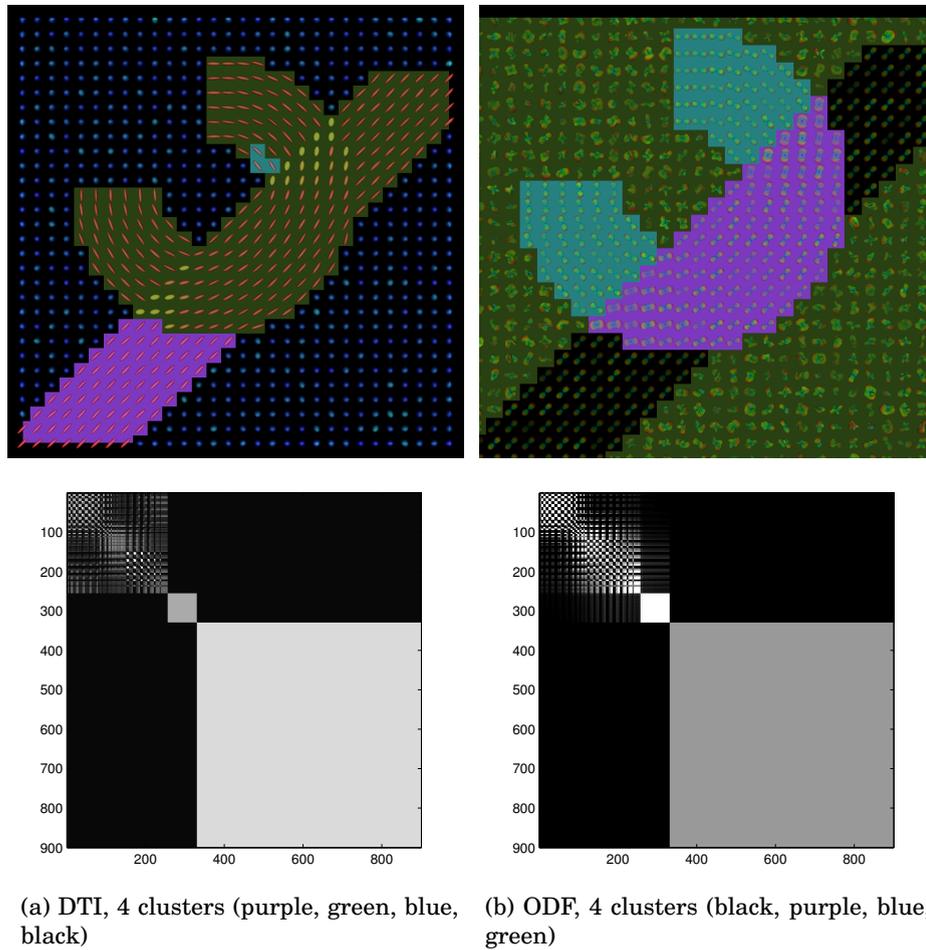


Figure 5.5: Clustering results in ODF and DT images, Only ODF show the correct clustering. In both cases the clustering result and the reordered similarity matrix are shown.

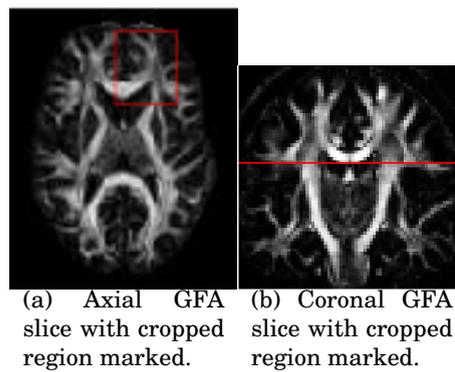


Figure 5.6: Generalized fractional anisotropy axial, 5.6a, and coronal, 5.6b slices in the real dataset.

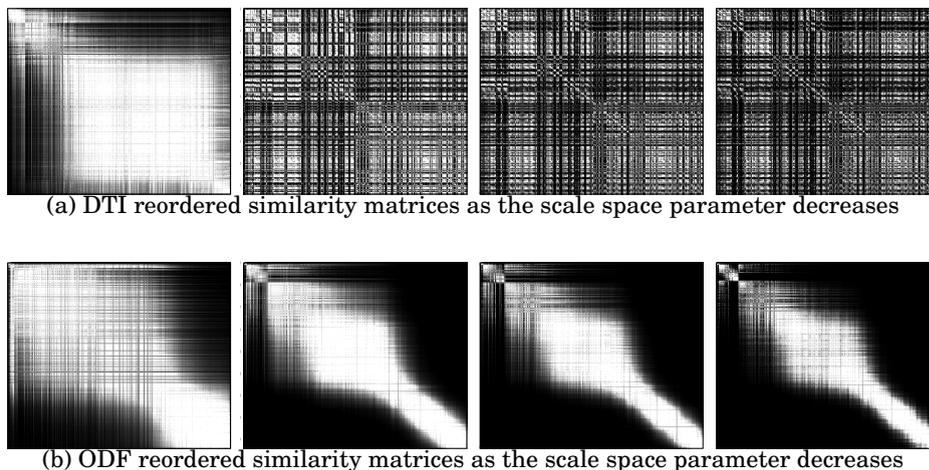


Figure 5.7: Plots of DTI and ODF similarity matrices of an axial cropped slice shown in figure 5.6. The matrices are reordered according to the second (Fiedler) eigenvector. The similarity matrices are shown in decreasing order of σ , which takes the values $\frac{1}{5}$, $\frac{1}{10}$, $\frac{1}{20}$ and $\frac{1}{40}$ of the quantity of elements to cluster. In the DTI case the decreasing on the scale parameter σ leads to a matrix with highly correlated elements that is very difficult to cluster. In the ODF case, the block structure is clear, hence better suited to apply a clustering algorithm.

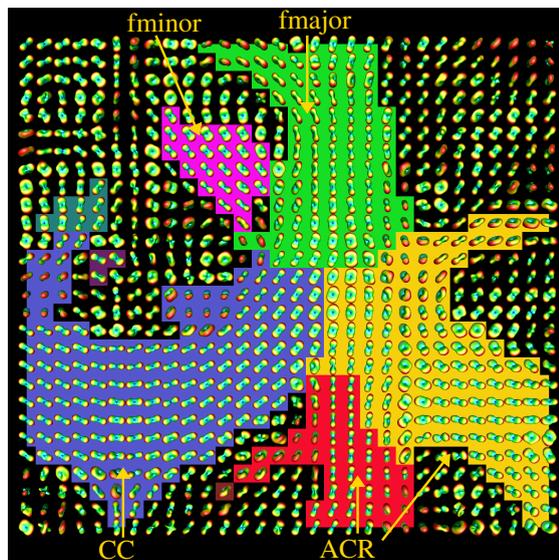
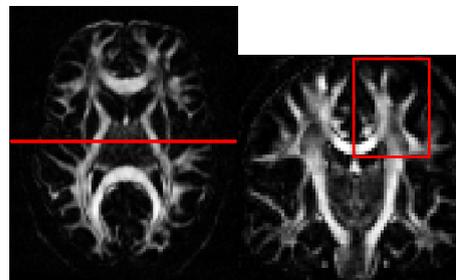
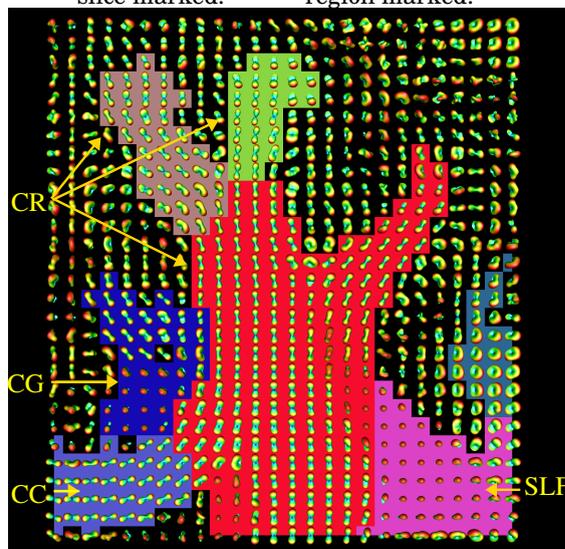


Figure 5.8: Our proposed algorithm is able to identify important white matter fibre bundles on an axial slice of a real dataset. The cropped axial slice shown in figure 5.6 has been segmented. In the labeled ODF visualization, each color represents one of the clusters found. The white matter labels are CC: Corpus Callosum, ACR: Anterior Corona Radiata, fmajor: Forceps Major and fminor: Forceps Minor.



(a) Axial GFA slice with coronal slice marked. (b) Coronal GFA slice with cropped region marked.



(c) Labeled ODF visualization of the cropped region.

Figure 5.9: Our proposed algorithm is able to identify important white matter fibre bundles on a coronal slice of a real dataset. Generalized fractional anisotropy axial, 5.9a, and coronal, 5.9b slices are shown. Labeled ODF visualization, 5.9c, each color represents one of the 7 clusters found. The white matter labels are CC: Corpus Callosum, CG: Cingulum, CR: Corona Radiata, SLF: Superior Longitudinal Fasciculus.

as the same clustering technique using QBI and DT images perform equivalently. However, in regions of fibre crossings the same algorithm using DT is clearly outperformed by its QBI version.

Having proved the performance of our algorithm in synthetic cases, we proceeded to deal with real data. We analysed the performance of our algorithm in sections of the human brain white matter where bundles cross in different directions. For this, we started by analysing the similarity matrix at different scales, we showed that due to the lack of orientational information in crossing areas, the DT-based version of our technique is not able to obtain clear sharp clusters segmenting white matter regions. Furthermore, as we performed this analysis on a similarity matrix obtained from the image and not on the clustering outcome, it is likely that this problem will be present in most clustering algorithms trying to perform the same task. However, the QBI-based version of our clustering algorithm, not only showed a clear block structure in its affinity matrices but it segmented successfully white matter tracts.

Limitations

The algorithm presented in this chapter has two main shortcomings. First, it requires to build a matrix that has as many entries as the squared number of voxels to cluster, after building this matrix its eigendecomposition must be performed. This results in a limitation on the use of the algorithm in large sections of an image. Even if recently there has been research aiming to reduce the computational and memory cost of these families of algorithms [Dhillon et al.: 2007, Fowlkes et al.: 2004], this adds new parameters to the algorithm which are not related to the clustering problem directly. The second limitation of this algorithm is due to the nature of the fibre crossings. Due to the fact that fibre crossing areas naturally belong to different white matter fibre bundles, an algorithm like the one proposed which produces a hard label for each segmented group of voxels is not appropriate. In fact, these areas should be labelled as belonging to several bundles at the same time, denoting that fuzzy-clustering algorithms, like the one presented by Awate et al. [2007] might better address the crossing-area problem if applied to QBI images.

5.5 SUMMARY AND CONCLUSION

We have presented an algorithm to perform Q-ball imaging segmentation of white matter fibre bundles. The proposed method combines state-of-the-art HARDI reconstruction and state-of-the-art spectral clustering techniques. Our algorithm is initialization-free and has only two parameters: a scale space parameter and the number of regions (clusters) to be found. Regarding this number of clusters parameter, we have proposed to estimate it automatically.

In segmenting the white matter bundles, we have introduced a spectral clustering

technique that does not require uniform sampling of the elements. The importance of this characteristic was shown through experiments. To do so, the similarity measure used incorporates an Euclidean distance measure between the spherical harmonic coefficients describing the Q-Ball ODFs as well as the spatial location distance between ODFs. The similarity measure and the metric induced in the embedded space are then used to cluster Q-Ball ODF images into multi-label segmentation representing the fibre bundles. Spectral embedding has already been applied to dMRI [Ziyan et al.: 2006]. However, to our knowledge, this is the first work using the Diffusion Maps that avoids the high dependence on element sampling. It is also the first work attempting Q-Ball ODFs.

We have illustrated that the ODFs are the desirable elements to use for clustering in the white matter because the classical DT model is limited in regions of fibre crossings. The ODF is even more attractive because of the recent analytical spherical harmonic solution to the ODF reconstruction [Anderson: 2005, Descoteaux et al.: 2007a, Hess et al.: 2006]. The analytical solution is in fact as fast as a standard DT least-square estimation. In this work, we believe that we have used the state-of-the-art ODF reconstruction method [Descoteaux et al.: 2007a], which is regularized, robust and very simple to implement. Note that some improvements have been proposed in order to estimate a solid angle-based ODF [Aganj et al.: 2009, Tristán-Vega et al.: 2009].

The spectral embedding performed by the Diffusion Maps technique is at the heart of our segmentation algorithm. Whereas other spectral embedding techniques have a tendency to produce artifacts in the presence of different sampling characteristics within a cluster, the technique used in this work greatly reduces this tendency by performing the simple linear algebra calculation shown in algorithm 4.2.

Spectral embedding techniques produce a representation of the embedded data based on element-to-element affinities. This leads to the fundamental issue: how to choose the similarity measure? It is a challenge to find a measure that incorporates similarities between elements as well as the spatial location difference between elements. For similarities between elements, we chose the Euclidean distance between spherical harmonic coefficients describing the ODFs. This approach is simple and very efficient because it allows to process the ODFs directly on the SH coefficients. The Euclidean distance has also been used successfully in a level set segmentation framework [Descoteaux and Deriche: 2007] and it would be interesting to compare our spectral clustering approach against it. For spatial location difference, we chose Markovian Relaxation in order to be consistent with the graph theoretical representation of the Diffusion Maps technique. Although this way of representing the distance involves an artificial elimination of all the non-neighboring relations of the ODF elements in the similarity matrix and an adjustment of the diagonal elements, we believe that the resulting similarity relations better represent the similarity better. The similarity of two neighboring elements at the beginning of the Markovian

Relaxation algorithm is represented by a function of the Euclidean distance between them. This similarity can be interpreted as the probability that a random walker has of going from the first element to the second. The similarity of two elements at the end of the relaxation is the probability of a random walker starting from one element and reaching the second in a certain number of steps.

The final step of our algorithm is k-means clustering. We believe that there is room for improvement in this last part of the algorithm. In the first place, the k-means algorithm needs an explicit number of clusters to find. This can be heuristically determined by analysing the eigenvalue plot or the reordered similarity matrix structure, as shown in this work. However, an automatic method that could find the number of clusters would considerably improve the algorithm. In the second place, the k-means algorithm and its variants, for instance, k-medians, k-medioids, search for isotropic clusters in the embedding space (see chapter 3) are able to perform clustering on convex structures. This could also improve the last clustering phase of our algorithm

Finally, in order to analyse the importance of the difference between our Diffusion Maps algorithm and the widely used N-Cuts, we used synthetic simulations. In these simulations, we generated a synthetic image with a single cluster within which the sampling of the elements changed. We showed that when this sampling changes, the N-cuts algorithm produces artefacts while our Diffusion Maps method does not. As uniform sampling within a cluster is a difficult property to guarantee in the white matter fibre bundles, our Diffusion Maps method is better suited for this task.

Part III

Gaussian Processes Modelling of the White Matter Fibres

Overview

In the past years, there has been an increase interest in clinical studies and applications involving DMRI-based fibre tracking . As a consequence, there is a growing demand for an objective mathematical framework to perform quantitative analysis of white matter fibre bundles. Part III of this thesis presents such a novel mathematical framework that facilitates mathematical operations between tracts. Examples of these operations are: a measure to compare fibres based on their spatial overlapping; the combination of fibre tracts into a bundle; and a quantification of the probability that a tract or voxel belongs to a bundle. Based on these operations, we developed an algorithm for automated *in vivo* dissection of white matter bundles. This algorithm performs unsupervised atlas-based clustering without requiring manual initialization or an *a priori* knowledge of the number of clusters. Also, using our novel framework we developed two clinical applications: straightening of the spinal cord in Spinal cord MRI (SC-MRI); and group statistics to find differences between healthy controls and schizophrenic subjects.

The mathematical framework introduced in this chapter has been published in [Wassermann et al. \[2010a\]](#) and [Wassermann et al. \[2009\]](#), its application to spinal cord MRI straightening in [Wassermann et al. \[2010c\]](#) and [Wassermann et al. \[2010b\]](#) and its application to statistics on white matter structures in [Wassermann et al. \[2010d\]](#).

Contributions of This Part:

- Definition of a mathematical framework for white matter bundles which facilitates statistical analyses. This framework incorporates spatial and diffusivity information and spans a metric space between bundles.
- Development of an algorithm to perform *in vivo* white matter fibre bundle dissection based on tractography.
- Development of an algorithm to select the most representative, or prototype, fibre of a bundle. Application to spinal cord straightening.
- Formulation of usual group statistical analyses in terms of our mathematical framework in order to find biomarkers for schizophrenia.

Introduction

In the previous part we have aimed to segment a DMRI image into its white matter bundles. Even if we had some success in this task, the integration of spatial and diffusivity information constitutes a challenge. Perhaps the voxel is too localized in order to be analysed on its own or as a group with the goal of identifying white matter structures. Having this said, we direct our research towards the analysis of tractography results. Diffusion MRI-based tractography provides a more global anatomical interpretation of the diffusivity on each voxel. By tracing ensembles of axonal pathways at a sub-voxel resolution it also provides a solution to some of the problems arising from partial voluming. Moreover, as seen in chapter 2, there is extensive proof of the reproducibility and anatomical coherence of tractography-obtained white matter fibre tracts.

The clustering of different fibre tracts into an anatomically coherent bundle, like the CC or the CST, is a challenging task for several reasons. In the first place, as seen in figure 5.10, axons composing a bundle can diverge from it connecting cortical and subcortical areas. This renders approaches that quantify similarity among white matter fibres using the whole fibre instead of analysing partial overlaps like shape statistics or rigid transformations [Veltkamp: 2001] unsuited for the clustering task. Take for instance the cingulum bundle, whose constituent fibres only partially overlap among themselves, with many diverging to innervate the cortex, as we have seen in chapter 2 and we shown in figure 5.10. These divergent fibres can have quite different shapes, calling into question the utility of shape-based metrics for subsequent tensor statistics. Even if current streamline tractography techniques are reproducible [Wakana et al.: 2007], fibre tracts obtained through tractography do not recover the whole underlying axonal trajectory [Lenglet et al.: 2009]. This behaviour can be observed clearly in complex bundle configurations like crossings or fannings [Savadjiev et al.: 2008], the crossing area of the CC and the CST for example [Wiegell et al.: 2000]. In order to overcome this problem, streamline tracking techniques that are more sensitive to complex bundle configurations have been developed [Qazi et al.: 2009, Descoteaux et al.: 2009a]. These produce better local approximations of axonal distribution by exploiting more complex models of the underlying water diffusion in a voxel [Tuch: 2004, Peled et al.: 2006, Descoteaux et al.: 2007a]. Still, results are far from being reliable after the streamline tracking pro-

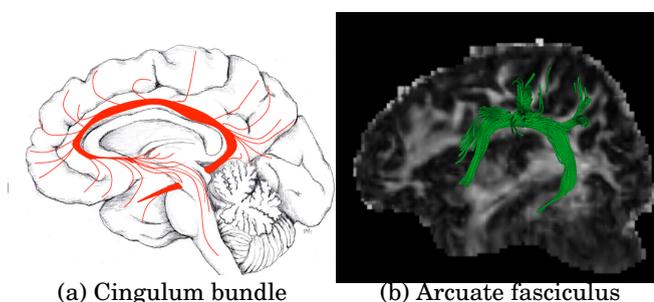


Figure 5.10: Axons enter and leave anatomical bundles. Image (a) reproduced from [O'Donnell \[2006\]](#) with permission of Jimmy Fallon, UCI. Image (b) manually selected Arcuate fasciculus after full brain tractography.

cedure traverses a region with complex bundle configuration. In clustering tracts into a bundle, a common workaround for this problem consists of seeding all over the brain and performing a dense whole brain tractography. This algorithm produces fibre tracts that could later be grouped through clustering techniques [[O'Donnell and Westin: 2007](#)]. This highlights the need for a similarity metric that can quantify the closeness of two fibres or the degree to which a tract belongs to a bundle; both of these are challenging problems, critical to automatic bundle identification.

Quantifying fibre similarity is a fundamental part of fibre clustering that has been addressed in different ways. Recent works [[Batchelor et al.: 2006](#), [Corouge et al.: 2006](#), [Leemans et al.: 2006](#)] quantify fibre similarity with different flavours of shape statistics. However, partial overlapping of fibres is not taken into account as a similarity feature. Thus, the previous approaches are unsuited for automatic classification of fibres in the brain. There is a separate set of works [[Ding et al.: 2003](#), [O'Donnell and Westin: 2007](#), [Wassermann and Deriche: 2008](#), [Maddah et al.: 2008a](#)], which uses different clustering algorithms based on the Hausdorff or Chamfer distances among the sequence of points parametrizing each fibre tract. This family of similarity metrics deals with sets of points instead of curves, hence they discard continuity or directionality information. Moreover, similarity tends to decrease very fast in cases of partial overlapping, failing to include fibres diverging from the bundle in the correspondent cluster. In particular, [Ding et al. \[2003\]](#) only analyse fibres whose seed points are spatially close together. This is not suited for a whole brain analysis because different fibre seed points from the same bundle may have been scattered all over the white matter. This seeding technique is frequently used in order to overcome limitations of streamline tracking in regions with complex bundle configurations [[O'Donnell and Westin: 2007](#)]. Manifold learning techniques are used by [O'Donnell and Westin \[2007\]](#) and [Wassermann and Deriche \[2008\]](#) to generalize these distances from small sets of similar fibres to a bigger more diverse set of fibres. These approaches embed the fibres into Euclidean or topological spaces that can be handled more easily.

[O'Donnell and Westin \[2007\]](#) start by generating an atlas of white matter fibres,

fibres from new subjects are then classified according to this atlas. Even though these automatically grouped bundles are anatomically coherent, the process to generate the atlas requires heavy user interaction and fine parameter tuning. The level of manual interaction needed renders the approach difficult to reproduce. [Wassermann and Deriche \[2008\]](#) use a publicly available anatomical atlas in conjunction with the fibre similarity metric. This work requires a smaller number of parameters, nevertheless situations of partial fibre overlapping generate non-anatomically coherent bundles. This strategy has proved to be useful for single individuals but lacks the necessary parameter stability needed for group studies: the parameters of the algorithm must be fine-tuned individually for each subject in order to obtain the same white matter bundle. [Maddah et al. \[2008a\]](#) enhances the Hausdorff similarity with Mahalanobis distance between fibre points. In order to handle partial overlapping, an ad-hoc penalty term is added to this distance. This approach requires user initialization, by selecting a fibre which is known to be in the desired bundle. Their subsequent work [[Maddah et al.: 2008b](#)] incorporated atlas information to increase accuracy, however an initial fibre representing each bundle is still required. Recently, [Neji et al. \[2009\]](#) and [Durrleman et al. \[2009\]](#) proposed mathematical frameworks to perform statistical analysis of white matter fibre bundles, however the main idea of these metrics still does not take in account partial fibre overlapping as their predecessors, putting into question their usability for clustering of human brain's white matter fibres. From all the presented approaches, only [O'Donnell and Westin \[2007\]](#) succeed in the task of semi-automated classification of the whole ensemble of white matter fibres, however this is achieved with a great deal of user interaction and parameters tuning.

Once bundles have been found, quantitative analysis can be performed. We described in detail some of these studies in chapter 2 but we briefly go over the literature again. Analysis of DMRI-based tractography is useful to monitor pathological conditions [[Ciccarelli et al.: 2008](#), [Kubicki et al.: 2007](#)]. Most of the works performing bundle statistics [[Goodlett et al.: 2009](#), [Hua et al.: 2008](#), [O'Donnell et al.: 2007](#), [Maddah et al.: 2008a](#)] rely on the use of medial representations for bundles. These representations are only appropriate for bundles which can be modelled as convex envelopes. Thus, their methodology is not entirely appropriate at the extremes of the bundles, an area where the axons fan-out innervating cortical or subcortical structures. Other works [[Oh et al.: 2007](#)] use a mesh approach over ROIs and perform statistics on the surface. However, automated transition from a set of fibres to the mesh is unclear. There is recent evidence that tract probability maps can be used in order to perform bundle-oriented statistics in diffusion MRI [[Hua et al.: 2008](#)] and histological images [[Bürgel et al.: 2006](#)]. However, in these two approaches the process to obtain the tract probability maps must be performed manually by experts. Overall, statistical models of white matter bundles which rely on medial representations are insufficient. Moreover, for models which are more appropriate for a wider

spectra of bundles, the transition from automatically obtained bundles to these statistical models is not straightforward.

OUR CONTRIBUTION

In this part of the thesis, we address the important problem of developing a mathematical framework for the quantitative analysis of fibre bundles. This research area has recently become very active, with the aim of facilitating subsequent clustering and group-based statistical analyses on the bundles. These statistical analyses aim to detect changes in the white matter structures. These changes, as we have seen in chapter 2, can be evident, as in the case of brain stroke that severs tracts, or subtle, as in the case of neuropsychiatric disorders, such as multiple sclerosis which can manifest as changes in the tracts.

The mathematical framework we propose in this part aims to set the foundation for statistical analysis of white matter fibre bundles. Particularly, by providing important operations among them: combination of white matter fibres into a bundle; quantification of bundle similarity based on their overlap in space; and the mapping of a point in space to its probability of belonging to a bundle, the tract probability map of the bundle. In order to, produce a mathematical formulation of these operations, white matter fibres and bundles are represented as *blurred indicator functions*. These functions are then parametrically represented as Gaussian Processes. This provides a setting to formulate the previously mentioned operations efficiently on the parameter space of these processes. Moreover, the Gaussian Processes are a suited tool to represent white matter bundles integrating spatial and diffusion tensor information. Hence, we have a set of operations to analyse white matter bundles, and we are able to provide an efficient implementation in terms of Gaussian processes.

Finally, we use this framework to develop fibre-analysis applications: a method for automated clustering of fibres into anatomical bundles like the arcuate fasciculus; an algorithm to select the prototype fibre of a bundle in the same vein as [Corouge et al. \[2006\]](#), [Maddah et al. \[2008a\]](#) and [O'Donnell et al. \[2009\]](#); and a technique to map bundles to tract probability maps, enabling tract-based statistics on the cerebral white matter like the ones used to characterize neuropathology effects by [Pagani et al. \[2005\]](#), [Lin et al. \[2007\]](#) and [Hua et al. \[2008\]](#).

The following chapters are as follows: First we review the Gaussian processes as a mathematical model to infer functions from samples and to define operations among bundles which are useful to perform statistical analysis. Then, we use these Gaussian processes in order to develop a model for the fibres and a tool to perform automatic identification of the human brain's white matter structures. Finally, we develop a prototype fibre selection algorithm and we use it to straighten spinal cord MRI images.

GAUSSIAN PROCESSES

Map drawing and spatial statistics are disciplines that have been around for a long time. Every traveler needs to orient himself and every farmer looks for different ways to manage his crops and animals, distribute them on the terrain and quantify the results. However, the necessity of reducing these maps to numbers and analyse this synthetic representation is fairly recent. The advent of new technologies, like remote sensing, capable of numerically quantify terrain characteristics, electroencephalography and magneto-encephalography, capable of measuring brain activity on the surface of the scalp, represent different cases of modern scenarios where spatial data analysis is used in order to extract information from a function sampled from a surface. Recently, spatial statistics have been used in combination with functional Magnetic Resonance Imaging to assess the significance of apparent signal observed in noisy difference images [Worsley et al.: 1996].

6.1 NON-LINEAR REGRESSION WITH A BAYESIAN FLAVOUR

When dealing with samples of spatial data, two main tasks arise in the pre-processing step: smoothing and interpolation: given a set of samples from a function $y(\cdot)$ whose domain is a line, surface or volume, we want to be able to add a reasonable set of hypotheses and infer, up to a certain error, the value on the whole domain. More formally, given N data points $X = \{\mathbf{x}_i\}_{i=1}^N$ sampled from a domain $D \subset \mathbb{R}$ and its corresponding sampling of the function values $Y = \{y_i = y(\mathbf{x}_i)\}_{i=1}^N$, we want to be able to predict $y_{N+1} = y(\mathbf{x}_{N+1})$ for any given $\mathbf{x}_{N+1} \in D$. Using the Bayesian paradigm, we formulate the inference of $y(\cdot)$ from the given samples as a posterior probability distribution:

$$\mathbb{P}\{y(\cdot)|Y, X\} = \frac{\mathbb{P}\{Y|y(\cdot), X\} \mathbb{P}\{y(\cdot)\}}{\mathbb{P}\{Y|X\}} \quad (6.1)$$

where the factor $\mathbb{P}\{Y|y(\cdot), X\}$ states the probability of the function values being Y when the chosen function is $y(\cdot)$ and the domain samples are X ; the factor $\mathbb{P}\{y(\cdot)\}$ states the probability of the function $y(\cdot)$ abiding to some prior on the function characteristics, for instance, smoothness and the dividend $\mathbb{P}\{Y|X\}$ accounts for the reliability on the function values Y given the domain points X , hence being able to represent

noise or uncertainty in the measurements.

6.1.1 A one-dimensional case

As an example, let us state a simple problem. Given samples of an one-dimensional function and the prior condition that it is least wiggly as it can. We want to find a function that closely fits the sampled function values and is smooth up to its second derivative. A function which minimizes the functional

$$M(y(\cdot)) = \sum_{i=1}^N (y(x_i) - y_i)^2 + \alpha \int dx [y''(x)]^2. \quad (6.2)$$

where α is a tuning parameter giving more importance to the sample approximation or to the smoothness hypothesis. The previous formulation can be restated into the Bayesian paradigm by identifying the second term of $M(y(\cdot))$ as the *prior* on the function $y(\cdot)$. Thereby, we state that $y(\cdot)$ is a more probable representation of the function we're looking for if it is less wiggly, and we represent this hypothesis probabilistically as

$$\ln(\mathbb{P}\{y(\cdot)\}) \propto -\alpha \int dx [y''(x)]^2. \quad (6.3)$$

Having the probabilistic representation, we are able to show that this *prior* bounds $y(\cdot)$ to be characterized by a parametric probability distribution. In order to do this, first we take a set of finite samples on $X = \{x_i\}_{i=1}^N \subset \mathbb{R}$ and we rewrite the *prior* as a probability over a finite number of random variables:

$$\mathbb{P}\{\{y(x_i)\}_{i=1}^N\} \propto \exp\left(-\alpha \sum_{i=1}^N [y''(x_i)]^2\right). \quad (6.4)$$

Taking $\mathbf{y} = [y(x_i)]_{i=1}^N$ as a column vector and D as the linear operator mapping $y(\cdot)$ to its derivative, we write the previous equation in matrix form as

$$\mathbb{P}\{\mathbf{y}\} \propto \exp\left(-\alpha (\mathbf{D}^2 \mathbf{y})^T (\mathbf{D}^2 \mathbf{y})\right) = \exp\left(-\alpha \mathbf{y}^T [(\mathbf{D}^2)^T \mathbf{D}^2] \mathbf{y}\right). \quad (6.5)$$

This shows that $\mathbb{P}\{\mathbf{y}\}$ is proportional to the p.d.f. of a Gaussian distributed random variable in \mathbb{R}^N with a mean determined by the first term of equation 6.2 and covariance matrix

$$\mathbf{C} = [(\mathbf{D}^2)^T \mathbf{D}^2]^{-1} \quad (6.6)$$

From this, we infer that, under the prior stated in equation 6.3, for every sampling on \mathbb{R} of size $N \geq 1$, the probability of those samples coming from a function which is a candidate function for our problem follows a multivariate Gaussian p.d.f. . But a function is an infinite set of values, hence we need to extend the previous reasoning from a multivariate p.d.f. to a stochastic process¹. We do this by applying *Kol-*

1. Or stochastic random field

mogorov's Existence Theorem [Kolmogorov: 1956]. This theorem states that if every finite sample of $y(\cdot)$ follows a Gaussian p.d.f. , then the stochastic process representing the function $y(\cdot)$ follows Gaussian distribution or, for shorter, $y(\cdot)$ is a *Gaussian Process*(GP). Analogously to Gaussian multivariate distributions, Gaussian Processes can be characterized by their mean and covariance functions:

$$y^*(x) = \mathbb{E} \{y(x)\} = \int y(x) \mathbb{P} \{x\} dx \quad (6.7)$$

$$c(s, t) = \mathbb{E} \{(y(s) - y^*(s))(y(t) - y^*(t))\} = \iint y(s)y(t) \mathbb{P} \{s, t\} ds dt - y^*(s)y^*(t), \quad (6.8)$$

where, in order to be coherent with the previous analysis of the finite sample

$$c(x_i, x_j) = \mathbf{C}_{ij}.$$

Thus, as the only restriction on the parameters of the multivariate Gaussian p.d.f. defined in equations (6.5) and (6.6) is that \mathbf{C} must be a symmetric positive semi-definite matrix, the only restriction is that $c(s, t)$ is a symmetric positive semi-definite function in s and t .

Characterizing the covariance function

Now that we know that $y(\cdot)$ is characterized by its mean and covariance functions, we calculate the covariance function from the *prior* given on equation 6.3. For this we use the prior on a finite sample (equation 6.5), and the matrix form of the covariance (equation 6.6). We rewrite the latter as

$$\left[(\mathbf{D}^2)^T \mathbf{D}^2 \right] \mathbf{C} = \mathbf{D}^4 \mathbf{C} = \mathbf{I}. \quad (6.9)$$

Furthermore, if we rewrite the previous expression as

$$\sum_{k=1}^N [\mathbf{D}^4]_{ik} \mathbf{C}_{kj} = \delta_{ij} \quad (6.10)$$

and then reformulate it for the continuous case instead of the finite sampling X we obtain

$$\int_{\mathbb{R}} D^4(s, u) c(u, t) du = \delta(s - t), \quad (6.11)$$

we can observe that $c(\cdot, \cdot)$ is the Green function of the fourth derivative operator [Wahba: 1990]. When we formulated the smoothness hypothesis, we didn't make any statement on the smoothness which was related with the domain of the function. Consequently, the prior does not change along the domain or, we have what is called a *stationary prior*. Knowing that we simplify equation 6.11 by performing a change of variables, taking $d(s - t) := c(s, t)$ and $r := s - t$, and rewriting it for a single variable

r [Duchon: 1977]

$$\frac{\partial^4}{\partial^4 r} d(r) = \delta(r). \quad (6.12)$$

Then, the solution to the previous equation is

$$d(r) = \frac{1}{6}|r|^3 + a_3 r^3 + a_2 r^2 + a_1 r + a_0. \quad (6.13)$$

As we previously stated, we want $c(s, t)$ hence $d(r)$, to be the covariance function of the stochastic process modelling $y(\cdot)$. For this, we need $d(r)$ to be symmetric and positive-definite or, equivalently, it must abide to

$$d(s - t) = d(t - s), \quad d(0) > 0 \quad \text{and} \quad |d(r)| \geq d(0). \quad (6.14)$$

In order to meet the first condition, symmetry, we must eliminate all the odd-powered terms from equation 6.13. Moreover, the last of the three conditions is not generally possible in the infinite case, but it is if take a fixed value $R \in \mathbb{R}$, restrict r to $[-R, R]$ and we clamp the function and its derivative:

$$d(r) = 0 \quad \text{and} \quad \frac{\partial}{\partial r} d(r) = 0 \quad \text{when } r = R. \quad (6.15)$$

Solving the differential equation given in equation 6.13 with the conditions given in equations (6.14) and (6.15) [Duchon: 1977], we find that

$$d(r) = \begin{cases} \frac{1}{12} (2|r|^3 - 3Rr^2 + R^3) & |r| \leq R \\ 0 & |r| > R \end{cases} \quad (6.16)$$

and our *stationary* covariance function is

$$c(s, t) = d(s - t). \quad (6.17)$$

As this function depends only on the distance between s and t , we can call it a *radial* function. There is extensive bibliography on symmetric positive-definite radial functions and their use for inferring a function from samples, we refer the interested reader to Buhmann [2004], Wahba [1990], Wendland [1995], Buhmann [1998] and the references therein.

Finally, we can state that $y(\cdot)$ has a stochastic representation as a GP where parameters of the distribution are its mean and covariance functions:

$$y(\cdot) \sim \mathcal{GP}(y^*(\cdot), c(\cdot, \cdot)) \quad (6.18)$$

where $c(s, t) = d(s - t)$

Going back to the Bayesian setting for regression we stated in equation 6.1 and using the parametric model inferred for $y(\cdot)$, we can model the inference problem as the

posterior on the parameters of $y(\cdot)$:

$$\mathbb{P}\{y^*(\cdot), c(\cdot, \cdot) | Y, X\} = \frac{\mathbb{P}\{Y | y^*(\cdot), c(\cdot, \cdot), X\} \mathbb{P}\{y^*(\cdot), c(\cdot, \cdot)\}}{\mathbb{P}\{Y | X\}}. \quad (6.19)$$

Inferring arbitrary values of the mean and covariance functions

Up to this point, we know that we are looking for a function $y(\cdot)$, that we have samples of the values of this function, $Y = \{y_i\}_{i=1}^n \subset \mathbb{R}$, corresponding to a sampling $X = \{x_i\}_{i=1}^n \subset \mathbb{R}$ on the domain of the function. Also, we know that the function follows a GP and the formulation for its covariance function. Hence, in order to be able to compute the value of the function at an arbitrary point of the domain, x_{N+1} , we need to compute the values of the mean and covariance functions at that point.

In the Bayesian framework we presented, computing the mean and covariance functions for a given point translates into characterizing the p.d.f. of an arbitrary value of the function, y_{N+1} , given the sampling X, Y . This p.d.f. is usually called the *predictive distribution* and it is equivalent to calculating, given the samplings X, Y , the probability that y_{N+1} has of being a value of the function $y(\cdot)$ at point x_{N+1} in the domain which is not necessarily in the original sampling:

$$\mathbb{P}\{y_{N+1} | x_{N+1}, X, Y\}. \quad (6.20)$$

We characterize this p.d.f. using properties of the Gaussian Processes, in particular the one given in equation **B.3**. First, equivalently to a subset of components of a multivariate Gaussian, the joint distribution of a sample of $y(\cdot)$ over $X \subset D$ follows a multivariate Gaussian p.d.f. . Due to this, it is possible to use Bayes' rule to state

$$\mathbb{P}\{y_{N+1} | x_{N+1}, X, Y\} = \frac{\mathbb{P}\{y_{N+1}, Y | x_{N+1}, X\}}{\mathbb{P}\{Y | x_{N+1}, X\}}. \quad (6.21)$$

where, as a consequence of the previous argument, the two probabilities on the right hand side are multivariate Gaussian distribution. Hence, by notating X and Y in vector form as $\mathbf{x} = [x_1, \dots, x_N]^T$ and $\mathbf{y} = [y_1, \dots, y_N]^T$ we are able to define the p.d.f. of the denominator of equation **6.21**

$$Y | x_{N+1}, X = Y | X = \mathbf{y} | \mathbf{x} \sim \mathcal{G}(\mathbf{y}^*, \mathbf{C}_{\mathbf{y}\mathbf{y}}), \quad [\mathbf{C}_{\mathbf{y}\mathbf{y}}]_{ij} = c(x_i, x_j) \quad (6.22)$$

and of its numerator

$$\begin{bmatrix} \mathbf{y} \\ \dots \\ y_{N+1} \end{bmatrix} \sim \mathcal{G} \left(\begin{bmatrix} \mathbf{y}^* \\ \dots \\ y_{N+1}^* \end{bmatrix}, \begin{bmatrix} \mathbf{C}_{\mathbf{y}\mathbf{y}} & \mathbf{C}_{\mathbf{y}} \\ \dots & \dots \\ \mathbf{C}_{\mathbf{y}}^T & c_{x_{N+1}} \end{bmatrix} \right) \quad (6.23)$$

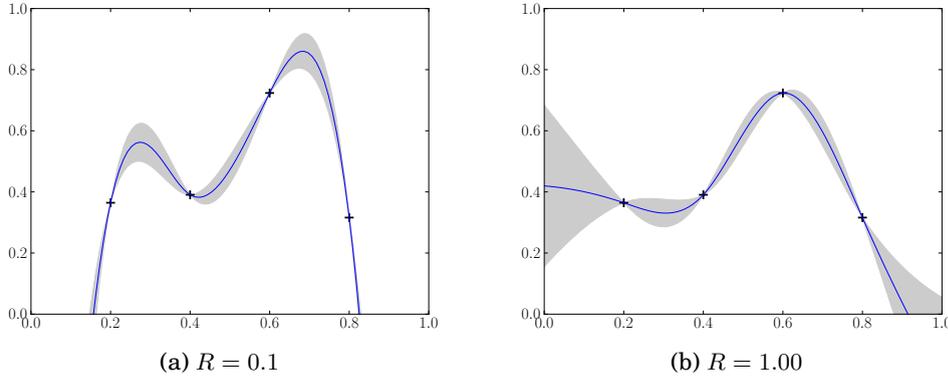


Figure 6.1: Gaussian Process-based one dimensional function regression. Four samples, noted as black crosses, are given as an input. The probabilistic *prior* on the function to infer is the one in equation 6.3. From the *prior*, a covariance function with a parameter R is derived, shown in equation 6.18. The mean function is inferred from equation 6.25 and plotted in blue. The shading represents the mean function plus-minus three times the standard deviation function, in equation 6.26, corresponding to the 3 standard deviations or the 99.7% confidence region.

where $\mathbf{C}_y = [c(x_1, x_{N+1}), \dots, c(x_N, x_{N+1})]^T$ and $c_{x_{N+1}} = c(x_{N+1}, x_{N+1})$. Then, by dividing the p.d.f. of the two Gaussian distributions of equation 6.21, as stated in equation B.3, and applying the partitioned inverse equation B.5, we can characterize the p.d.f. of y_{N+1} is the univariate Gaussian

$$y_{N+1} = y(x_{N+1}) \sim \mathcal{G}(y_{N+1}^*, \sigma_{N+1}^2), \quad (6.24)$$

where its mean and variance are inferred from the samplings X and Y as

$$y_{N+1}^* = y^*(x_{N+1}) = \mathbf{C}_y(x_{N+1})^T \mathbf{C}_{yy}^{-1} \mathbf{y}, \quad (6.25)$$

$$\sigma_{N+1}^2 = \sigma^2(x_{N+1}) = c_{x_{N+1}} - \mathbf{C}_y(x_{N+1})^T \mathbf{C}_{yy}^{-1} \mathbf{C}_y(x_{N+1}). \quad (6.26)$$

This characterization of y_{N+1} provides, in fact, an estimation of the function value at an arbitrary point: first, due to the previous reasoning, we know that y_{N+1} is the value of the function $y(\cdot)$ at x_{N+1} , $y_{N+1} = y(x_{N+1})$; second, that given x_{N+1} the most probable value of $y(x_{N+1})$ is y_{N+1}^* . In figure 6.1 we show the inferred function $y(\cdot)$, as a blue line, from a set of four samples, marked as black crosses. Moreover, we also use a shading in order to illustrate the 99.7% confidence region, or $y^*(x) \pm 3\sqrt{\sigma^2(x)}$. It can be seen in this illustration that, through GP-based regression, we were able to infer a continuous function which is parametrically represented by its mean and covariance functions.

Finally, it is simple to generalize the GP framework for Bayesian regression to higher dimensions, what is called Gaussian Random Fields. This is achieved by solving the integral equation 6.11 for \mathbb{R}^K , $K > 1$. For instance, solving it for the

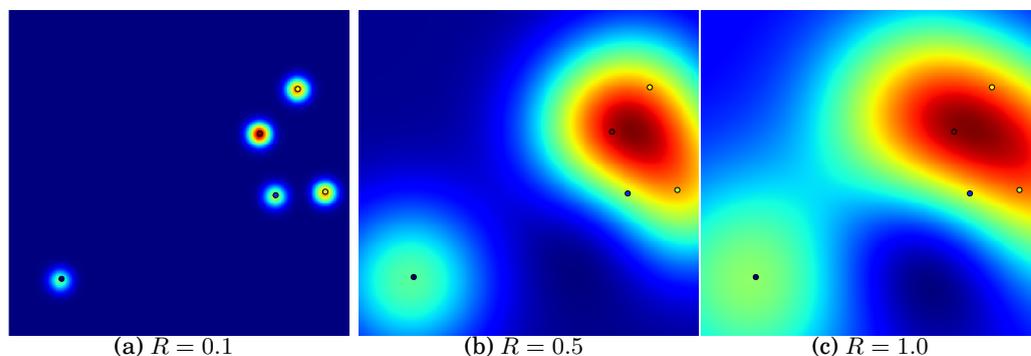


Figure 6.2: Gaussian Process-based two-dimensional function regression. Five samples, noted as dots colored by their value, are given as an input. The probabilistic *prior* on the function to infer is the one in equation 6.3. From the *prior*, a covariance function with a parameter R is derived, shown in equation 6.27. The mean function is inferred by means of equation 6.25 and shown in the same color code as the colored dots.

two-dimensional case leads to

$$c(\mathbf{s}, \mathbf{t}) = d(\|\mathbf{s} - \mathbf{t}\|),$$

$$\text{where } d(r) = \begin{cases} 2r^2 \log |r| - (1 + 2 \log(R))r^2 + R^2 & |r| \leq R \\ 0 & |r| > R \end{cases} \quad (6.27)$$

Then, the same regression procedure we used for the one-dimensional case can be applied. In figure 6.2 mean functions for a set of 5 samples and different values of R are shown.

In this section we have derived a *stationary* covariance function from a smoothness hypothesis. However, this is not the only covariance function available, the use of different covariance functions has been studied in different fields. Different flavours of these functions have emerged from a wide range of hypotheses and case studies. We refer the interested reader to reviews of its use in Machine Learning by [Rasmussen and Williams \[2006\]](#) and [MacKay \[1998\]](#) and in spatial statistics by [Cressie \[1991\]](#) and [Stein \[1999\]](#).

6.2 NON-STATIONARY PRIORS ON FUNCTION REGRESSION _____

While Gaussian Process-based models of function regression based on *stationary priors* have been successfully used in spatial statistics [[Stein: 1999](#)], in Machine Learning [[Rasmussen and Williams: 2006](#)] and in statistical analysis of brain functions [[Worsley et al.: 1996](#)], these approaches have the weakness of being global models. In these, the variability of the estimated process is the same throughout the domain because the covariance function is equal over its whole extension.

This lack of adaptability to variable or heterogeneous scenarios of the underlying function is of particular importance in every application field. Take for instance a mountain scenario, typical of the Andes mountains, where beside to rocky

and uneven mountains there are cliffs and plain valleys. Recently, spatial statistics researchers made progress in defining and constructing several kinds of non-stationary covariance functions [Paciorek and Schervish: 2006]. Covariance functions are the way to express *prior* information on the inferred function, therefore, from a Bayesian point of view, this is equivalent to proposing non-stationary *priors* on the function to be inferred. In order to retain the focus on the tools used in this work, we solely present the general way of constructing non-stationary covariance functions through convolving spatially varying kernel functions, a technique introduced by Higdon et al. [1999] and Fuentes and Smith [2001] and generalized by Paciorek and Schervish [2006]. We refer the interested reader to the works of Rasmussen and Williams [2006], MacKay [1998] and [Paciorek and Schervish: 2006] for reviews of non-stationary covariance functions and their applications to different research fields.

The main result proposed by Higdon et al. [1999] is that given a spatially evolving kernel $K(\cdot; x)$, the function

$$c^{NS}(\mathbf{s}, \mathbf{t}) = \int K(\mathbf{u}; \mathbf{s})K(\mathbf{u}; \mathbf{t})d\mathbf{u} \quad (6.28)$$

is the covariance function of a *non-stationary* Gaussian Process. The fact that $c^{NS}(\mathbf{s}, \mathbf{t})$ is symmetric is simple to verify, therefore we only need to prove that is positive semi-definite:

$$\sum_{i=1}^N \sum_{j=1}^N a_i a_j \int K(\mathbf{u}; \mathbf{x}_i)K(\mathbf{u}; \mathbf{x}_j)d\mathbf{u} = \quad (6.29)$$

$$\int \sum_{i=1}^N a_i K(\mathbf{u}; \mathbf{x}_i) \sum_{j=1}^N a_j K(\mathbf{u}; \mathbf{x}_j) d\mathbf{u} \quad (6.30)$$

$$\int \left(\sum_{i=1}^N a_i K(\mathbf{u}; \mathbf{x}_i) \right)^2 d\mathbf{u} \geq 0. \quad (6.31)$$

Then, if $w(x)$ is a white noise Gaussian Process

$$w(\cdot) \sim \mathcal{GP}(w^*(\cdot), \delta(\mathbf{s}, \mathbf{t})), \quad (6.32)$$

the stochastic process

$$y(x) = \int K(\mathbf{u}; x)w(\mathbf{u})d\mathbf{u} \quad (6.33)$$

has a Gaussian Process p.d.f. and its covariance function is:

$$\mathbb{E}\{(y(\mathbf{s}) - y^*(\mathbf{s}))(y(\mathbf{t}) - y^*(\mathbf{t}))\} = \mathbb{E}\{y(\mathbf{s})y(\mathbf{t})\} - y^*(\mathbf{s})y^*(\mathbf{t}) \quad (6.34)$$

where [Higdon et al.: 1999, Fuentes and Smith: 2001]

$$\mathbb{E}\{y(\mathbf{s})y(\mathbf{t})\} = \int \int K(\mathbf{u}; \mathbf{s})K(\mathbf{u}'; \mathbf{t})\delta(\mathbf{u}, \mathbf{u}')d\mathbf{u}d\mathbf{u}' = \int K(\mathbf{u}; \mathbf{s})K(\mathbf{u}; \mathbf{t})d\mathbf{u}. \quad (6.35)$$

This family of covariance functions is able to model spatial functions where the correlation between two data points changes with their location.

6.2.1 A Gaussian example

In order to illustrate an application of non-stationary Gaussian Process, we show two inferred functions from the same set of samples on a domain D shown in figure 6.3a. Each inference uses a different covariance function:

Stationary inference The covariance function for this scenario is given by previously known isotropic Gaussian p.d.f. :

$$c^S(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right),$$

with $\sigma = 0.5$. Figure 6.3b shows, for set of points in the domain, the area comprehended within one standard deviation of the corresponding covariance function in green.

Non-stationary inference The covariance function for this scenario is given by an anisotropic Gaussian p.d.f. which depends on the position:

$$c^{NS}(\mathbf{x}, \mathbf{y}; \mathbf{z}) = \frac{1}{2\pi\sqrt{|A(\mathbf{z})|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{y})^T A^{-1}(\mathbf{z})(\mathbf{x} - \mathbf{y})\right),$$

where $A(\mathbf{z} = (z_1, z_2)) = R(-\text{atan2}(z_1, z_2)) \begin{pmatrix} .5 & 0 \\ 0 & .1 \end{pmatrix}$ and $R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$.

Figure 6.3c shows, for set of points Z in the domain, the area comprehended within one standard deviation of the corresponding covariance function, $A(\cdot, \cdot; \mathbf{z})$, $\mathbf{z} \in Z$ in green.

Then, we perform the inference process described in section 6.1 using the two covariance functions. For each point in D , we used the set of samples shown in figure 6.3a and the above described covariance functions to infer the mean function of the corresponding GP as described in subsection 6.1.1. The resulting function for the stationary case can be observed in figure 6.3d where it is noticeable how the boundary exhibits a wiggled behaviour due to the choice of the covariance function. In the non-stationary case, figure 6.3e, as the covariance function converges to a diagonal at a location close to the boundary shown by the samples (figure 6.3a) it can be observed that the inferred function (figure 6.3e) shows a straight line corresponding to the boundary.

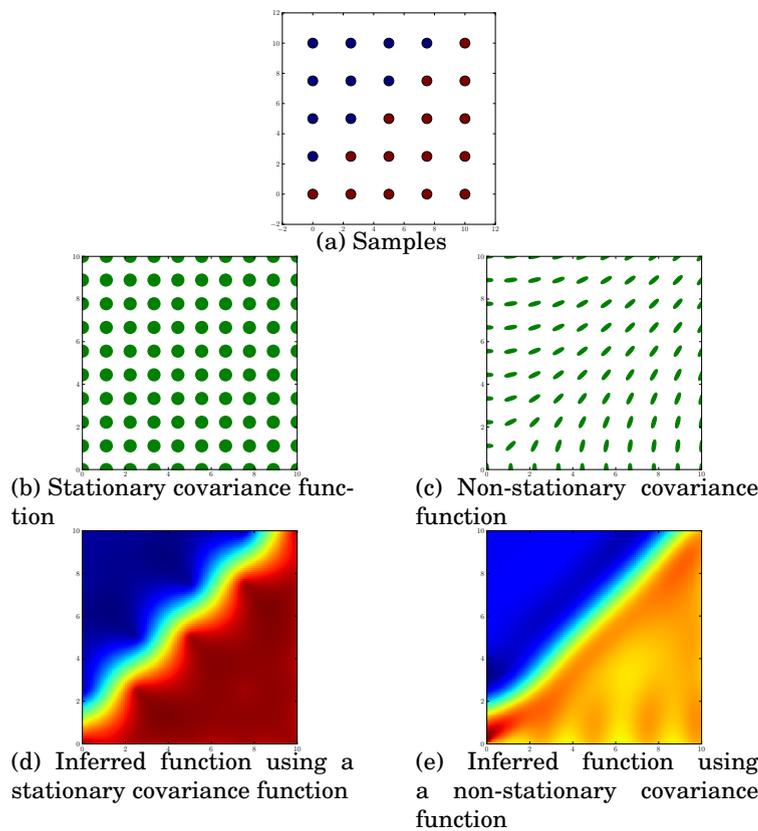


Figure 6.3: Inference of a function given by a set of samples using stationary and non-stationary covariance functions. Samples are shown in subfigure (a), where it can be noted a boundary at the diagonal. In order to illustrate the stationary and non-stationary covariance functions, the support of the covariance functions at a set of sampled points is represented by green circles and ellipsoids, shown in subfigures (b) and (c). When the prior is stationary, the circles are equal all over the domain (subfigure (b)) and the result of the inference process (subfigure (d)) shows artifacts at the boundary. When the prior is non-stationary, converging to a diagonal line at the centre, it can be observed that the sharpness of boundary is preserved, subfigure (e).

Overall, we have shown that non-stationary covariance functions constitute a flexible tool to include a rich set of hypotheses in GP-based regression. Moreover, we have also shown that using a non-stationary covariance function does not require any change in the inference framework, provided that we know it in advance.

6.3 FUNCTION INNER PRODUCT SPACE

In this work, the main goal of using the Gaussian Process framework is to be able to perform statistical analysis on functions which are inferred from samples. The main advantage of the GP representation of the functions is its usability as a proper space to perform statistics: we can build a vector space, and furnish it with an inner product operation, based on the parameter space of the Gaussian Processes. The combination of the vector space with an inner product operation results in an inner product space. Within this space, we can interpolate functions and quantify their similarity, two fundamental operations for performing statistical analysis. In this

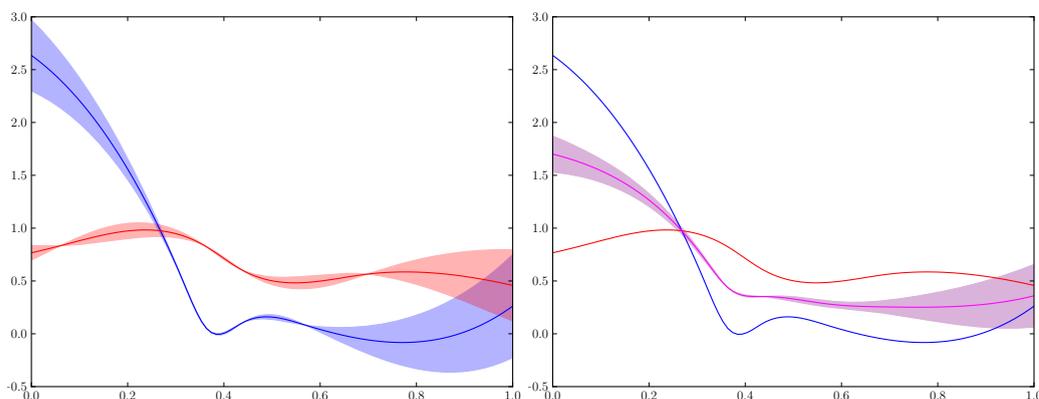


Figure 6.4: Average of two Gaussian Processes. On the left, in red and blue, we show the mean function and 99.7% confidence interval of the Gaussian Processes of two functions inferred from samples. On the right, in purple, is the Gaussian Process of the mean function and its 99.7% confidence interval calculated by equation 6.44

section, we show how to build a vector space of functions based on the GP.

The main results of this section are the linear combination between two functions represented as Gaussian processes

$$\alpha y_1(\cdot) + \beta y_2(\cdot) \sim \mathcal{GP}(\alpha y_1^*(\cdot) + \beta y_2^*(\cdot), \alpha^2 c_1(\cdot, \cdot) + \beta^2 c_2(\cdot, \cdot)) \quad (6.36)$$

and an inner product operation

$$\langle y_1(\cdot), y_2(\cdot) \rangle := \int y_1^*(x) y_2^*(x) dx, . \quad (6.37)$$

As direct consequence of the previous characteristics, a norm is induced by the inner product:

$$\|y(\cdot)\| := \langle y(\cdot), y(\cdot) \rangle \quad (6.38)$$

These two operations combined with the linear combination form a vector space of functions represented as Gaussian Processes. This constitutes a comfortable space to operate and perform statistical analysis of functions inferred from a finite number of samples and *a priori* information about their smoothness.

6.3.1 Vector Space

We begin by formulating the vector space properties among functions, meaning scalar multiplication and addition, in terms of Gaussian processes. It is easy to prove that the GP representation of the functions can represent these operations in parameter space as the linear combination of two multivariate Gaussian r.v. is also representable in terms of their multivariate distribution: being this the case for any finite sample of the GP, due to Kolmogorov's consistency theorem, they are valid for

the GP itself. Then, given two functions represented as GPs

$$y_1(\cdot) \sim \mathcal{GP}(y_1^*(\cdot), c_1(\cdot, \cdot)) \quad (6.39)$$

$$y_2(\cdot) \sim \mathcal{GP}(y_2^*(\cdot), c_2(\cdot, \cdot)) \quad (6.40)$$

the following properties apply

$$\alpha y_1(\cdot) \sim \mathcal{GP}(\alpha y_1^*(\cdot), \alpha^2 c_1(\cdot, \cdot)) \quad (6.41)$$

$$y_1(\cdot) + y_2(\cdot) \sim \mathcal{GP}(y_1^*(\cdot) + y_2^*(\cdot), c_1(\cdot, \cdot) + c_2(\cdot, \cdot)). \quad (6.42)$$

The definition of these operations verifies, as with the multivariate Gaussian r.v., that functions represented as GPs constitute a vector space. Consequently, linear combination of functions in terms of the parameters of their corresponding GPs, shown in equation 6.36, is defined by combining the two previous rules. From the two operations defined above, it is straightforward to define linear interpolation

$$\begin{aligned} \alpha y_1(\cdot) + (1 - \alpha)y_2(\cdot) &\sim \mathcal{GP}(y_\alpha^*(\cdot), c_\alpha(\cdot, \cdot)), \quad \alpha \in [0, 1] \\ y_\alpha^*(\cdot) &= \alpha y_1^*(\cdot) + (1 - \alpha)y_2^*(\cdot), \\ c_\alpha(\cdot, \cdot) &= \alpha^2 c_1(\cdot, \cdot) + (1 - \alpha)^2 c_2(\cdot, \cdot) \end{aligned} \quad (6.43)$$

and averaging

$$\frac{1}{N} \sum_{i=1}^N y_i(\cdot) \sim \mathcal{GP} \left(\frac{1}{N} \sum_{i=1}^N y_i^*(\cdot), \frac{1}{N^2} \sum_{i=1}^N c_i(\cdot, \cdot) \right). \quad (6.44)$$

of functions. We illustrate the mean of two GPs in figure 6.4, and in figure 6.5 the interpolation of the mean functions of two GP is shown. Thus, the functions represented as GPs form a vector space and the operations of this vector space are defined in terms of the parameters of the GPs.

6.3.2 Inner product space

Now that we have stated that the functions represented as GPs are a vector space, and that we can operate on the parameter space their distributions, we proceed to furnish this vector space with an inner product operation.

We define an inner product operation that accounts for the L_2 norm between functions. Moreover, in order to reduce the computational complexity of this operation, we define a deterministic inner product operation. To provide such operation, we need to make a decision about the value of $y(x)$ defined in equation 6.25. This is done by employing *decision theory* [De Groot: 2004]: to make a decision about the value of $y(x)$ at x , we use a point-like prediction, $y^+(x)$. This prediction is taken in order to minimise the error in the squared norm induced by our inner product (

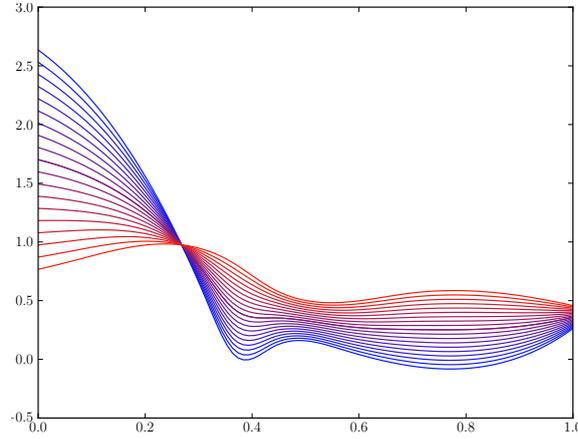


Figure 6.5: Interpolation of two Gaussian Processes. We show the mean function of the Gaussian Process corresponding to two functions inferred from samples in red and blue. Between both of them, the mean function of several Gaussian Processes obtained by equation 6.43

equation 6.38)

$$\operatorname{argmin}_{y^+(x)} \int (y^+(x) - y(x))^2 \mathbb{P}\{y(x)|\mathbf{y}, \mathbf{t}, x\} dy(x) = y^*(x). \quad (6.45)$$

Thus letting the mean value of $y(x)$, $y^*(x)$, be an appropriate estimator of the value of $y(x)$ at x . Then, as $y^*(x)$ is square integrable due to its definition in equation 6.25,

$$\langle y_1(\cdot), y_2(\cdot) \rangle := \int y_1^*(\cdot) y_2^*(\cdot) dx$$

is an inner product [Schmidt: 1908]. Furthermore, it can be easily computed by replacing equation 6.25 in the previous formula:

$$\begin{aligned} \langle y_1(\cdot), y_2(\cdot) \rangle &:= \int (C_{\mathbf{y}_1}(x)^T C_{\mathbf{y}_1 \mathbf{y}_1}^{-1} \mathbf{y}_1)^T (C_{\mathbf{y}_2}(x)^T C_{\mathbf{y}_2 \mathbf{y}_2}^{-1} \mathbf{y}_2) dx \\ &= \int (\mathbf{y}_1^T (C_{\mathbf{y}_1 \mathbf{y}_1}^{-1})^T) C_{\mathbf{y}_1}(x) C_{\mathbf{y}_2}(x)^T (C_{\mathbf{y}_2 \mathbf{y}_2}^{-1} \mathbf{y}_2) dx. \end{aligned}$$

Moreover, we can simplify the previous expression using that $C_{\mathbf{y}_1}(x)$ and $C_{\mathbf{y}_2}(x)$ are the only vectors depending on x :

$$\langle y_1(\cdot), y_2(\cdot) \rangle := \mathbf{y}_1^T (C_{\mathbf{y}_1 \mathbf{y}_1}^{-1})^T \left(\int C_{\mathbf{y}_1}(x) C_{\mathbf{y}_2}(x)^T dx \right) C_{\mathbf{y}_2 \mathbf{y}_2}^{-1} \mathbf{y}_2 \quad (6.46)$$

where

$$\left[\int C_{\mathbf{y}_1}(x) C_{\mathbf{y}_2}(x)^T dx \right]_{ij} = \int c([\mathbf{y}_1]_i, x) c([\mathbf{y}_2]_j, x) dx. \quad (6.47)$$

This final expression of the inner product, equation 6.46, can be calculated analytically whenever there is a closed form solution for equation 6.47.

This not only completes our definition of an inner product space for functions

represented as GPs but also shows how function interpolation and the calculation of the inner product between two of them can be carried on within the parameter space of the GPs.

6.4 SUMMARY

In this chapter, we recalled the basics of Gaussian Processes in order to state function interpolation within the Bayesian Paradigm. In the following part of this thesis, Gaussian Processes will play a fundamental role to develop a statistical framework for white matter fibre tracts calculated from tractography.

GAUSSIAN PROCESS FRAMEWORK FOR WHITE MATTER BUNDLES

The goal of this chapter is to introduce a novel mathematical framework for performing statistical analysis of fibre tracts and bundles. Our model includes diffusion information and relates the bundles with an ROI in the volume, mapping every voxel to degree of membership to the bundle, the bundle's *blurred indicator function*. It provides a similarity measure for fibre bundles and fibre tracts, which are considered as single-fibre bundles. Its linear combination operation between fibre bundles seamlessly generates new bundle configurations and allows for the volume-based statistics of fibre bundles. In addition, the similarity measure handles cases of partial fibre overlap naturally. The previously mentioned characteristics facilitate statistical analysis and classification/clustering tasks. Finally, we present a clustering application based on our mathematical framework and on anatomical information in the shape of a volumetric atlas. The output of this application is a set of automatically obtained white matter bundles like the Arcuate Fasciculus or the Cingulum. For each bundle, we are also able to produce an ROI which maps every voxel to its probability of belonging to the bundle, referred in previous work as *tract probability map* [Hua et al.: 2008, Bürgel et al.: 2006]. This map is an appropriate tool to perform bundle-based statistics on the cerebral white matter. We validate this clustering algorithm by applying it to 68 healthy subjects and then performing statistical analysis on the results using our mathematical framework.

7.1 INTUITION

The first step in designing our framework is to provide a mathematical model for white matter fibres. Each fibre tract \mathcal{F} is modeled as a *blurred indicator function* $y_{\mathcal{F}} : \mathbf{p} \in \mathbb{R}^3 \rightarrow \mathbb{R}$. This *blurred indicator function* has a maximal level set which corresponds to \mathcal{F} . Moreover, $y_{\mathcal{F}}$ is blurred in accordance with the information provided by the underlying diffusion tensor field, i.e. along the fibre direction and not across it. This is illustrated on figure 7.1.

Then, we need to govern the properties of our fibres/functions while subse-

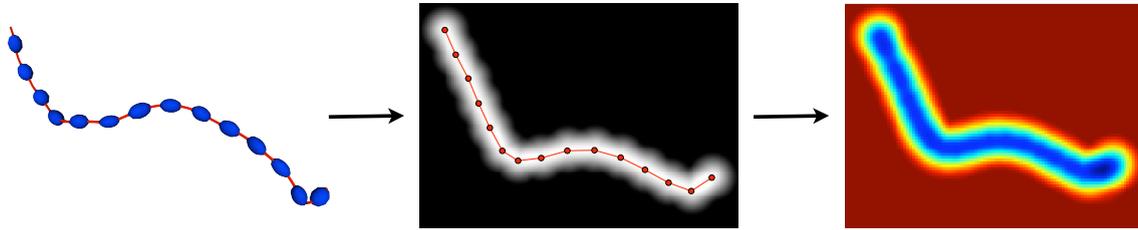


Figure 7.1: Gaussian Process representation of the fibre: We model each fibre as a blurred indicator function. Each fibre obtained by tractography is constituted as a sample of points over a line, which is shown in red line on the left side of the figure, and its corresponding diffusion tensor field, indicated by the blue tensors over the red fibre. The sampled points are then used to generate a blurred indicator function with a maximal level set which corresponds to the fibre (central image of the figure). Then, the fibre is blurred in accordance with the information provided by the underlying diffusion tensor field. The resulting representation, on the right side of the figure, is a smooth indicator function blurred along the fibre direction.

quently simplifying a certain number of sophisticated operations between the fibres like similarity quantification and combination into bundles. Gaussian Processes (GP) [Seeger: 2004], as we show in chapter 7, provide the right framework to integrate spatial and diffusion tensor MRI information for $y_{\mathcal{F}}$ and to perform these operations between the fibres. A Gaussian Process can be seen as a generalization of the classical Gaussian probability distribution to describe properties of functions and not only properties of random variables such as scalars or vectors. More precisely, we model the *blurred indicator function* $y_{\mathcal{F}}(\mathbf{p})$ by the GP

$$y_{\mathcal{F}}(\mathbf{p}) \sim \mathcal{GP}(y_{\mathcal{F}}^*(\mathbf{p}), c_{\mathcal{F}}(\mathbf{p}, \mathbf{p}')), \quad (7.1)$$

where the mean function $y_{\mathcal{F}}^*(\mathbf{p})$ and covariance function $c_{\mathcal{F}}(\mathbf{p}, \mathbf{p}')$ are the parameters of this stochastic process. These two functions are inferred from the tractography of each fibre. That is, from the sequence of points $\mathbf{f} = \{\mathbf{f}_1, \dots, \mathbf{f}_{|\mathcal{F}|}\}$ estimated by tractography of the anatomical bundle \mathcal{F} and from the corresponding sampling on its diffusion tensor field $\Sigma(\mathbf{f}_1), \dots, \Sigma(\mathbf{f}_{|\mathcal{F}|})$. The inference process of the GP corresponding to a white matter fibre is developed in subsection 7.2.1. It is important to point out that such framework provides also adequate computational tractability. Through closed form operations on the parameters of the GPs, it allows us to measure bundle similarity with partial fibre overlap without relying on point correspondences and to combine different fibres into a bundle by simply averaging the fibres' GPs. Now that we have a Gaussian Process representation of white matter fibre bundles, we can express similarity, combination and *tract probability maps* in terms of this representation.

Firstly, we implement a similarity measure between bundles. Our measure quantifies the overlapping as of two bundles as illustrated in figure 7.3. This is done with an inner product operation developed in detail in subsection 7.2.3. If the mean function representing a bundle \mathcal{F} , $y_{\mathcal{F}}^*(\mathbf{p})$, is square integrable and has finite support,

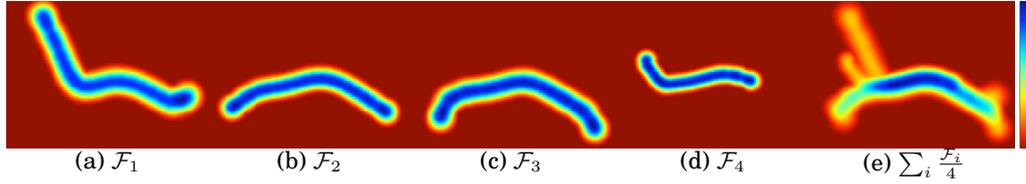


Figure 7.2: Mean indicator function for four fibre tracts (a-d) and mean indicator function for the bundle formed by averaging them according to our framework (e). Blue color means that the bundle is more likely to cross that voxel while red color means it is not likely that the bundle traverses that voxel. Fibres were manually selected from a full brain tractography and belong to the Cingulate Cortex section of the Cingulum (CgC).

the inner product between two bundles \mathcal{F} and \mathcal{F}' is defined as

$$\langle \mathcal{F}, \mathcal{F}' \rangle := \int_{\mathbb{R}^3} y_{\mathcal{F}}^*(\mathbf{p}) y_{\mathcal{F}'}^*(\mathbf{p}) d\mathbf{p}, \quad (7.2)$$

along with its induced norm

$$\|\mathcal{F}\|^2 := \langle \mathcal{F}, \mathcal{F} \rangle. \quad (7.3)$$

Moreover, we can define a similarity measure, bounded by 1 when \mathcal{F} and \mathcal{F}' are exactly the same and by 0 when there is no intersection, as the inner product normalized by its induced norm,

$$\langle \mathcal{F}, \mathcal{F}' \rangle_N := \frac{\langle \mathcal{F}, \mathcal{F}' \rangle}{\|\mathcal{F}\| \|\mathcal{F}'\|}. \quad (7.4)$$

Examples of these two similarity measures are shown in figure 7.3.

Using our framework, we are able to combine N fibres into a bundle by simply averaging the GPs corresponding to these fibres. The GP which corresponds to the indicator function of a fibre bundle \mathcal{B} , is obtained through the mean Gaussian Processes of single fibres or smaller bundles composing it,

$$y_{\mathcal{B}}(\mathbf{p}) = \frac{1}{N} \sum_{i=1}^N y_{\mathcal{F}_i}(\mathbf{p}) \sim \mathcal{GP}(y_{\mathcal{B}}^*(\mathbf{p}); c_{\mathcal{B}}(\mathbf{p}, \mathbf{p}')), \quad (7.5)$$

where $y_{\mathcal{B}}^*(\mathbf{p}) = \frac{1}{N} \sum_{i=1}^N y_{\mathcal{F}_i}^*(\mathbf{p})$ and $c_{\mathcal{B}}(\mathbf{p}, \mathbf{p}') = \frac{1}{N^2} \sum_{i=1}^N c_{\mathcal{F}_i}(\mathbf{p}, \mathbf{p}')$.

An example of this is shown in figure 7.2, where blurred indicator functions for four fibre tracts and the obtained function for the bundle combining them can be seen.

It is worth noting that the combination of fibres bundles, which is a linear combination operation, and the similarity measure, which is an inner product operation, constitute an inner product space. As we will show later, this space is an appropriate support to perform statistical analyses such as clustering of fibre bundles.

We mentioned previously the importance for bundle statistics of the *tract probability map*, the probability that a point \mathbf{p} in \mathbb{R}^3 is contained in a bundle \mathcal{B} , $\mathbb{P}\{\mathbf{p} \in \mathcal{B}\}$. As we show in subsection 7.2.4, having calculated the GP for a bundle \mathcal{B} , $y_{\mathcal{B}}(\cdot)$, by

	FF & UNC	FF & FF	FF & FF	CgC & CgC
$y_{\mathcal{F}}^*(\mathbf{p})$				
$y_{\mathcal{F}'}^*(\mathbf{p})$				
$y_{\mathcal{F}}^*(\mathbf{p})y_{\mathcal{F}'}^*(\mathbf{p})$				
$\langle \mathcal{F}, \mathcal{F}' \rangle$	4.26	115.61	48.03	2089.
$\frac{\langle \mathcal{F}, \mathcal{F}' \rangle}{\ \mathcal{F}\ \ \mathcal{F}'\ }$	0.093	0.800	0.474	0.600

Figure 7.3: Examples of the product of *blurred indicator functions* for different fibre pairs, the value of our inner product operation $\langle \mathcal{F}, \mathcal{F}' \rangle$, defined in equation 7.2, and of our inner product normalized by its natural norm, $\|\mathcal{F}\| = \sqrt{\langle \mathcal{F}, \mathcal{F} \rangle}$. Inner product quantifies the overlapping of *blurred indicator functions*. A larger inner product means that fibres are more similar and relates to the volume of the overlapping. The inner product normalized by its norm quantifies similarity ranging from 0 when overlapping is null to 1 when the two fibres are identical. The compared fibres have been extracted from different anatomical tracts of a diffusion MRI image, the Frontal Forceps (FF), the Uncinate Fasciculus (UNC) and the Cingulate Cortex section of the Cingulum (CgC).

means of equation 7.5, we can define the *tract probability map* from the parameters of the GP as

$$\mathbb{P}\{\mathbf{p} \in \mathcal{B}\} \propto \frac{1}{2\sqrt{\pi(h^2 + \sigma_{\mathcal{B}}^2(\mathbf{p}))}}, \quad (7.6)$$

where h is a parameter that diffuses the *tract probability map* in space and $\sigma_{\mathcal{B}}^2(\mathbf{p})$ is calculated from the parameters of $y_{\mathcal{B}}(\cdot)$. In order to illustrate the *tract probability map* calculated from a GP for a bundle, color-coded surfaces and a probability map over an FA image for anatomical bundles are shown in figure 7.4.

Up to this point, we have introduced our GP-based framework for white matter fibre bundles and its three main operations: combination of fibres into a bundle, similarity quantification and calculation of the *tract probability map*. Thus, in this section we provide all the right and necessary tools to perform a quantitative statistical analysis of white matter fibre bundles. In the following sections, we use these tools in order to perform automatic white matter bundle identification by means of white matter fibre clustering and we assess the quality of this clustering by performing inter-subject statistical analysis of white matter fibre bundles using our framework.

7.2 GAUSSIAN PROCESS FRAMEWORK

Gaussian Processes (GP) have been introduced in chapter 6. In designing a model for the white matter fibre bundles, we use GPs to produce a parametric representation. Particularly, we take advantage of its capability to incorporate different types

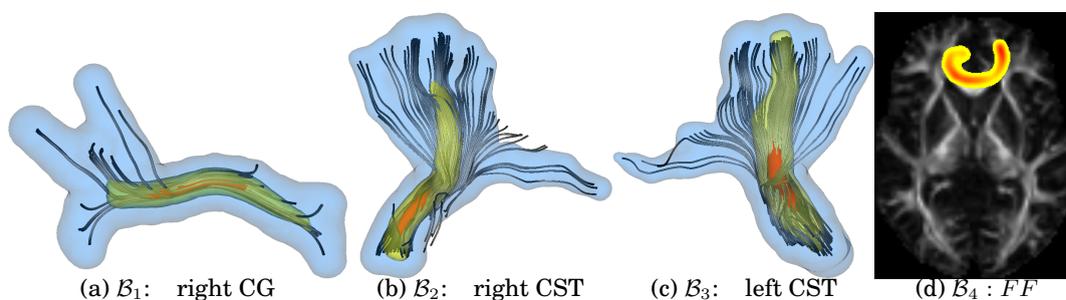


Figure 7.4: Iso-probability surfaces for manually selected fibre bundles, fig. (a-c), and tract probability map over FA for an automatically obtained bundle, fig. (d), see chapter 3. The probability at each voxel is calculated using equation 7.6. Color code for fig. (a-c) is as follows, Blue: $\mathbb{P}\{\mathbf{p} \in \mathcal{B}_i\} = .01$, Yellow: $\mathbb{P}\{\mathbf{p} \in \mathcal{B}_i\} = .2$, Red: $\mathbb{P}\{\mathbf{p} \in \mathcal{B}_i\} = .6$.

of hypotheses into the model. This representation provides a simple way to linearly combine fibres into bundles and to measure similarity through an inner product operation which we develop in subsection 7.2.3. Moreover, it provides a natural method to calculate *tract probability maps*, developed in subsection 7.2.4.

7.2.1 Single fibres as Gaussian Processes

Our parametric representation of a single fibre bundle is based on two hypotheses. Firstly, *smoothness*: due to the fact that fibre bundles in the brain do not have sharp angles [Basser et al.: 2000], we consider that the least wiggled trajectory joining the sample points of a fibre represents that fibre in a most probable manner. Secondly, *diffusion associated blurring*: The decay of the *blurred indicator functions* at each point on the bundle can be modeled using the water diffusion profile at that position. This relates the width of the bundle with the two smallest eigenvalues and their eigenvectors.

Using these two hypotheses, we write $y_{\mathcal{F}}(\mathbf{p})$, the GP for the indicator function of fibre \mathcal{F} , $y(\mathbf{p})$ for clarity, as the combination of two other GPs, $y_s(\mathbf{p})$ and $y_d(\mathbf{p})$. The process $y_s(\mathbf{p})$ represents the *smoothness* of the trajectory in space, its parameters are inferred from the point sequence obtained through the tractography of a fibre: $\mathbf{f} = \{\mathbf{f}_1, \dots, \mathbf{f}_{|\mathcal{F}|}\} \subset \mathbb{R}^3$. The process $y_d(\mathbf{p})$ represents the diffusion information, adds a variability to the fibre at \mathbf{p} using full *diffusion information* and it is inferred from the tensor field over the fibre $\Sigma(\mathbf{f}_1), \dots, \Sigma(\mathbf{f}_{|\mathcal{F}|})$. We now show that $y_s(\mathbf{p})$ and $y_d(\mathbf{p})$ can be modelled as GPs by characterizing them through covariance functions.

Smoothness

In chapter 6 we showed that GPs are a fit parametrical model for smooth functions. Particularly, this is the case of *blurred indicator functions*. Having stated that functions abiding to a smoothness constraint are representable by GPs in the aforementioned chapter, we proceed to fully characterize the GP correspond-

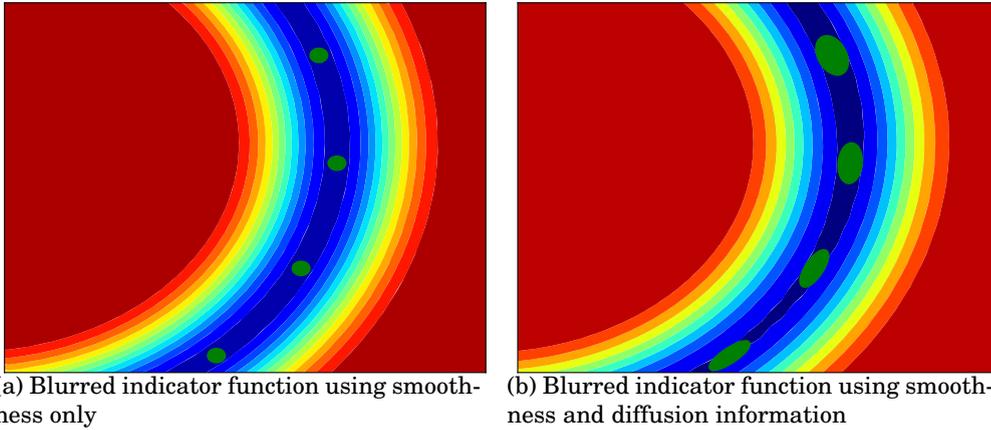


Figure 7.5: Segment of two *blurred indicator functions* for the same fibre. On the left, the green dots represent the samples of the fibre f , only the smoothness-related Gaussian Process is used to generate the *blurred indicator function*. On the right smoothness and diffusion associated blurring, the green ellipses denote the diffusion tensors. It can be seen how, the decay from the fibre to the background is even along the fibre on the left, while on the right depends on the directional diffusion intensity represented by the diffusion tensors.

ing to the *blurred indicator functions*, which consists of deriving the appropriate covariance function. This amounts to find the solution of equation 6.11 for the three-dimensional case: the solution for the previously mentioned equation inside a sphere in \mathbb{R}^3 of radius R is

$$c_s(\mathbf{p}, \mathbf{p}') := \psi(\|\mathbf{p} - \mathbf{p}'\|), \psi(r) = \begin{cases} 2|r|^3 - 3Rr^2 + R^3 & r \leq R \\ 0 & r > R \end{cases}, \quad (7.7)$$

where the constants have been chosen such that $c_s(\mathbf{p}, \mathbf{p}')$ is a positive semi-definite symmetric function.

Up to this point, we have fully characterized the probabilistic space of functions which describes the *blurred indicator function* for a smooth trajectory as the family of GP whose covariance function is given by equation 7.7. In subsection 7.2.2 we show how this is useful to infer the value of $y_s(\cdot)$ at an arbitrary point in space from a finite set of samples. The result of this inference process is shown on figure 7.5a.

Diffusion associated blurring

Until now, we have only used the points which constitute a fibre, to build the *blurred indicator function*. This function has a maximal value for points with high probability of being at the fibre and decays to 0, meaning no probability of belonging to the fibre, at an even speed across the fibre as seen in figure 7.5a. Within this section, we develop the means to perform the blurring using full tensor information. This means that on every direction, the intensity of the blurring depends on the intensity of the diffusion of water molecules. This diffusion is modelled by the diffusion tensors associated with the fibre which were obtained through diffusion

tensor MRI. This will enhance the blurring of the *blurred indicator function* along the fibre and relate the decay across it to the width of the diffusion tensors as depicted in figure 7.5b.

We now represent the diffusion information of a point in the trajectory by a *blurring* GP at that point. The usual practice in GP literature is to perform isotropic blurring on every sampled point of the function [MacKay: 1998]. In our case, we use anisotropic blurring at every given point sampled from fibre by means of a second covariance function based on diffusion tensor MRI information. This covariance function is built with convolution kernels as detailed in subsection 6.1.1. More precisely, the anisotropic that we apply at a point of the *blurred indicator function* at a point \mathbf{p} uses diffusion tensor information. For this, we need to encode DT information into a *blurring* kernel $k(\cdot; \mathbf{p})$: we take the kernel as the probability of a particle going from \mathbf{p} to \mathbf{w} in a time τ in terms of the Diffusion Tensor $\Sigma(\mathbf{p})$ [Basser et al.: 1994b],

$$k(\mathbf{w}; \mathbf{p}) = \mathbb{P}\{\mathbf{w}|\mathbf{p}, \tau, \Sigma(\mathbf{p})\} = \frac{1}{\sqrt{(4\pi\tau)^3|\Sigma(\mathbf{p})|}} \exp\left(-\frac{1}{4\tau}(\mathbf{w} - \mathbf{p})^T (\Sigma(\mathbf{p}))^{-1} (\mathbf{w} - \mathbf{p})\right). \quad (7.8)$$

Finally, by performing the integral in equation 6.28, we characterize the covariance function for the *anisotropic blurring process* as,

$$c_d(\mathbf{p}, \mathbf{p}') = \frac{1}{\sqrt{(4\pi\tau)^3|\Sigma(\mathbf{p})+\Sigma(\mathbf{p}')|}} \exp\left(-\frac{1}{4\tau}(\mathbf{p} - \mathbf{p}')^T (\Sigma(\mathbf{p}) + \Sigma(\mathbf{p}'))^{-1} (\mathbf{p} - \mathbf{p}')\right).$$

To conclude this section, having characterized the covariance function of the anisotropic blurring process, the blurring at each point is represented by a zero mean GP,

$$y_d(\mathbf{p}) \sim \mathcal{GP}(0, c_d(\mathbf{p}, \mathbf{p}')).$$

This GP, as seen on the following section, combined with the GP representing the smooth *blurred indicator function*, $y_s(\mathbf{p})$, produces a smooth function blurred in accordance to diffusion tensor MRI information as shown in figure 7.5b.

Gaussian Process Representation of a Fibre

We are now in position to write the GP formulation for the *blurred indicator function* of the fibre:

$$y(\mathbf{p}) \sim \mathcal{GP}\left(y^*(\mathbf{p}) = y_s^*(\mathbf{p}); c(\mathbf{p}, \mathbf{p}') = c_s(\mathbf{p}, \mathbf{p}') + c_d(\mathbf{p}, \mathbf{p}')\right). \quad (7.9)$$

Up to this point we have a Gaussian Process-based model for the white matter fibres. This model incorporates spatial and diffusion information. Moreover, within this model we can linearly combine fibres into bundles as show in equation 7.5 and

quantify similarity as shown in equation 7.2. In the remainder of this appendix we show how to characterize the value of the indicator function $y(\mathbf{p})$ at any test point $\mathbf{p} \in \mathbb{R}^3$, how to effectively calculate the similarity between two fibres and finally, how to calculate the *tract probability map* for a bundle.

7.2.2 Calculating the indicator function value distribution for a test point

We want to calculate a p.d.f. for the value of $y(\cdot)$ at a test point $\mathbf{p} \in \mathbb{R}^3$. This is a simple operation as we have a GP representation of the indicator function $y(\cdot)$ for a trajectory \mathcal{F} , see equation 7.9. The p.d.f. of $y(\mathbf{p})$, given the tracked point sequence $\mathbf{f} = \{\mathbf{f}_1, \dots, \mathbf{f}_{|\mathbf{f}|}\} \subset \mathbb{R}^3$ and its corresponding diffusion tensor field $\Sigma = \{\Sigma(\mathbf{f}_1), \dots, \Sigma(\mathbf{f}_{|\mathbf{f}|})\} \subset SPD(3)$ are characterized as Gaussian-distributed random variable. More precisely, $y(\mathbf{p})$ is the Gaussian distribution,

$$(y(\mathbf{p})|\mathbf{f}, \Sigma, \mathbf{p}) \sim \mathcal{G}(y^*(\mathbf{p}), \sigma^2(\mathbf{p})) \quad (7.10)$$

due to marginalization properties of the GPs [MacKay: 1998]. Setting $y(\mathbf{p})$ to take the constant value l when \mathbf{p} is a point that belongs to the fibre trajectory \mathcal{F} , the mean and covariance functions can be calculated in the following way [MacKay: 1998]:

$$y^*(\mathbf{p}) = S_{\mathbf{f}}(\mathbf{p})^T C_{\mathbf{ff}}^{-1} \mathbf{1}l \quad \sigma^2(\mathbf{p}) = c_s(\mathbf{p}, \mathbf{p}) - S_{\mathbf{f}}(\mathbf{p})^T C_{\mathbf{ff}}^{-1} S_{\mathbf{f}}(\mathbf{p}), \quad (7.11)$$

where $[S_{\mathbf{f}}(\mathbf{p})]_i = [c_s(\mathbf{f}_i, \mathbf{p})]_i$, $[C_{\mathbf{ff}}]_{ij} = [c(\mathbf{f}_i, \mathbf{f}_j)]_{ij}$ with $1 \leq i, j \leq |\mathbf{f}|$ and $\mathbf{1}$ is the vector with all ones; the functions $c_s(\cdot, \cdot)$ and $c(\cdot, \cdot)$ were defined in equation 7.7 and equation 7.9. This formulation is equivalent to “train” a Gaussian Process-based regression with values l at the sampled fibre points and 0 everywhere else. We set the parameter R such that we guarantee that the *blurred indicator function* is compact, as the maximal distance between two consecutive points in \mathbf{f} . The parameter τ modulates the scale of the diffusion associated covariance function by setting a diffusion time. We set it as the maximal time needed to traverse the maximal distance between two consecutive sampled points, R . More precisely, letting $\lambda_1(\Sigma)$ be the largest eigenvalue of the tensor Σ , the parameter is

$$\tau = \max_{i=1..|\mathbf{f}|} \frac{R}{\sqrt{\lambda_1(\Sigma(\mathbf{f}_i))}}$$

Examples of *blurred indicator functions* can be seen in figures 7.2, 7.3 and 7.5.

7.2.3 Inner Product Space of Fibre Bundles

Having the white matter bundles represented as GPs, similarity between two of them can be measured in terms of the parameters of their GPs using the inner product operation defined in subsection 6.3.2. Consequently, due to the fact that GPs

constitute an inner product space (subsection 6.3.2). An interesting outcome of the inner product space property of our framework is the simple calculation of similarity between fibre bundles. We are able to calculate this similarity among two bundles of white matter fibres, $\mathcal{B} = \{\mathcal{F}_1, \dots, \mathcal{F}_N\}$ and $\mathcal{B}' = \{\mathcal{F}'_1, \dots, \mathcal{F}'_M\}$,

$$\langle \mathcal{B}, \mathcal{B}' \rangle = \left\langle \frac{1}{N} \sum_{i=1}^N y_{\mathcal{F}_i}(\mathbf{p}), \frac{1}{M} \sum_{j=1}^M y_{\mathcal{F}'_j}(\mathbf{p}) \right\rangle,$$

by using the linearity and symmetry properties of the inner product operation. Then, the previous equation becomes

$$\langle \mathcal{B}, \mathcal{B}' \rangle = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M \langle y_{\mathcal{F}_i}(\mathbf{p}), y_{\mathcal{F}'_j}(\mathbf{p}) \rangle. \quad (7.12)$$

This provides a quick and simple way to calculate the similarity between to fibre bundles using the fibre-to-fibre similarities. In chapter 7 we use this advantage to perform clustering of fibre bundles.

7.2.4 Calculation of the Tract Probability Map

Having characterized $y(\cdot)$ as a GP in subsection 7.2.1, we can express the probability that a point \mathbf{p} in \mathbb{R}^3 is contained in a bundle \mathcal{F} , this map is called the *tract probability map*. We do this by calculating the probability of $y(\mathbf{p}) = l$ or, equivalently the expected concentration of the random value $y(\mathbf{p})$ around l ,

$$\mathbb{P}\{\mathbf{p} \in \mathcal{F}\} := \mathbb{P}\{y(\mathbf{p}) = l | \mathbf{f}, \Sigma, \mathbf{p}\} \propto \mathbb{E}[\theta(y(\mathbf{p}) - l) | \mathbf{f}, \mathbf{t}, \mathbf{p}]. \quad (7.13)$$

where $\theta : \mathbb{R} \rightarrow [0, 1]$ is a symmetric kernel. To ease the equations and the computation time, we take $\theta(\cdot)$ as a Gaussian kernel, with standard deviation h ,

$$\theta(y(\mathbf{p}) - l) = \frac{1}{2\sqrt{\pi}h} \exp\left(-\left(\frac{y(\mathbf{p})-l}{h}\right)^2\right).$$

Then, we calculate equation 7.13 as

$$\mathbb{E}[\theta(y(\mathbf{p}) - l) | \mathbf{f}, \mathbf{t}, \mathbf{p}] = \int \theta(y(\mathbf{p}) - l) \mathbb{P}\{y(\mathbf{p}) | \mathbf{f}, \mathbf{t}, \mathbf{p}\} dy(\mathbf{p})$$

which leads to

$$\mathbb{P}\{\mathbf{p} \in \mathcal{F}\} \propto \mathbb{E}[\theta(y(\mathbf{p}) - l) | \mathbf{f}, \mathbf{t}, \mathbf{p}] = \frac{1}{2\sqrt{\pi}(h^2 + \sigma^2(\mathbf{p}))}. \quad (7.14)$$

where h is a bandwidth parameter and $\sigma^2(\mathbf{p})$ is defined in equation 7.11. Then, the *tract probability map* for a bundle \mathcal{F} on a domain Ω , is calculated by evaluating $\mathbb{P}\{\mathbf{p} \in \mathcal{F}\}$ at every point $\mathbf{p} \in \Omega$. In order to illustrate the probabilistic map for a

Algorithm 7.1 Fibre bundle clustering algorithm

Inputs: a full brain tractography $F = \{\mathcal{F}_i\}$, with $1 \leq i \leq |F|$, calculate the set of stochastic processes representing each fibre $Y = \{y_{\mathcal{F}_i}(\mathbf{p})\}$.

Outputs: A dendrogram T .

- 1: Initialize the set of edges in the dendrogram: $T = \emptyset$
- 2: **while** there is a pair of different bundles $\mathcal{B}, \mathcal{B}'$ in B , s.t. $\langle \mathcal{B}, \mathcal{B}' \rangle > 0$ **do**
- 3: Select two different bundles $\mathcal{B}, \mathcal{B}'$ such that $\langle \mathcal{B}, \mathcal{B}' \rangle = \max_{\mathcal{C}, \mathcal{C}' \in B} \langle \mathcal{C}, \mathcal{C}' \rangle$.
- 4: Remove the bundles \mathcal{B} and \mathcal{B}' from B and add the bundle $\{\mathcal{B} + \mathcal{B}'\}$.
- 5: Add the edge $(\mathcal{B}, \mathcal{B}')$ to the dendrogram T .
- 6: **end while**
- 7: **return** T

bundle, color-coded surfaces and a probability map over an FA image for anatomical bundles are shown in figure 7.4. These *tract probability maps* are highly similar to the hand-obtained ones by Hua et al. [2008] on DTI images and Bürgel et al. [2006] by means of chemical staining. These two previous works have shown that these *tract probability maps* are an appropriate tool to perform statistics on white matter fibre bundles.

7.3 CLUSTERING ALGORITHM

Taking advantage of the mathematical framework for fibre bundles presented in chapter 7, we propose a stochastic process-based agglomerative clustering algorithm which is a variant of the algorithm presented in chapter 3. This algorithm is executed over a full brain tractography, a set of densely sampled fibres from the whole white matter. Once executed, our algorithm generates a dendrogram, a tree structure where each joint is a candidate cluster, this is illustrated in figure 7.6. Then, this dendrogram can be interactively explored in order to choose the desired granularity of the clustering without reprocessing the data. Being a hierarchical agglomerative algorithm it has several desirable properties: To begin with, convergence is guaranteed by the finite number of elements to cluster. Next, the number of clusters does not have to be known *a priori*. Also, it handles the clustering of outliers. These are incorporated to clusters during the late stages of the clustering algorithm if at all, which makes easy to identify them [Jain et al.: 1999]. All of these characteristics make our clustering algorithm effective and robust in order to classify white matter fibres from a full brain tractography into anatomically coherent bundles.

Our clustering algorithm applied to a full brain tractography is described in algorithm 7.1. The output of this algorithm is a dendrogram T , more precisely a set of trees where joint edge represents the joining of two bundles such as the one shown in figure 7.6.

A main advantage of our framework within this clustering algorithm is that the most important operation in step 2, the inner product among bundles described in chapter 7, is fast and simple to compute as we show in subsection 7.2.4. Once we

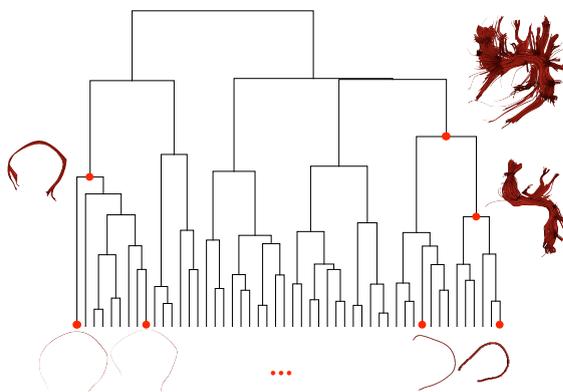


Figure 7.6: Illustration of our clustering algorithm defined in chapter 3. The algorithm starts from all the single-fibre bundles obtained by a full brain tractography, shown at the bottom of the figure, and joins them into multi-fibre bundles according to our similarity measure (equation 7.2). This generates a tree structure called dendrogram. Finally every joint is a candidate cluster. Sample clusters are shown on the side of the dendrogram and their positions on the dendrogram are marked with red dots.

have calculated the matrix of inner products for every pair of fibres in the full brain tractography \mathcal{F} , the algorithm works by simply performing linear operations on the rows of this matrix.

7.3.1 Tract Querying: Automatic cluster selection based on anatomical knowledge

Once our clustering algorithm generates the dendrogram for a full brain tractography, chapter 3, the main problem is how to select the joints in the dendrogram so that they are anatomically correct clusters. In order to do this, we introduce a query system based on volumetric information, the result of this query will be a cluster selected from joints of the dendrogram. In this work we use a publicly available atlas which has a parcellation of the brain gyri on the grey and white matter [Wakana et al.: 2004] as anatomically-aware volumetric information. Then, an anatomical query is defined by a set of grey or white matter regions that the tract must traverse, for example, the *Inferio Fronto Occipital* tract must connect the *inferio-frontal gyrus* and the *medial-occipital gyrus*. Images of the white matter atlas parcellation and queries for various tracts are shown in figure 7.7. After setting an anatomical query Q traversing K labeled regions on the atlas, $Q = \{r_1, \dots, r_K\}$, we use the *tract probability map* of each bundle, equation 7.6, to select the bundle \mathcal{B} on the dendrogram T with maximal joint probability of traversing all the regions:

$$\mathcal{B} = \operatorname{argmax}_{\mathcal{B}' \in T} \prod_{r \in Q} \int_{\mathbf{p} \in r} \mathbb{P} \{ \mathbf{p} \in \mathcal{B}' \} d\mathbf{p}. \quad (7.15)$$

The results of queries for 13 different white matter tracts, table 7.1, on 4 subjects are exhibited in figures 7.9 and 7.9 and *tract probability maps* of the mean bundles across

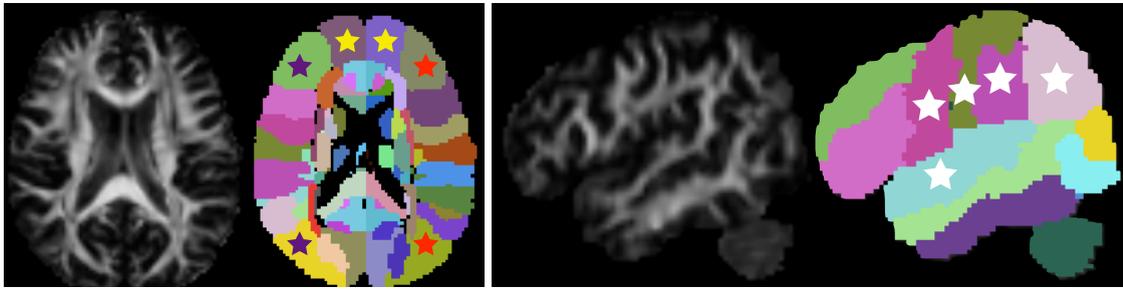


Figure 7.7: Illustration of anatomical queries based on volumetric information. A Fractional Anisotropy image and an image of a parcellation of the white matter are shown. The colored stars indicate the anatomical queries. On the left: Purple and red stars tag the inferio-frontal gyrus and middle-occipital gyrus on the left and right sides respectively, this query corresponds to the inferio-fronto-occipital fasciculus on the left and right hemispheres. Yellow stars tag the left and right middle-frontal-orbital gyrus, this query corresponds to the frontal forceps. On the right: white stars tag the pre and post central gyri, the angular, supra-marginal and superio-temporal gyri, this query corresponds to the arcuate fasciculus. The results of this queries on 4 subjects are exhibited in figures 7.9 and 7.9 and *tract probability maps* for the mean bundles for every subject are shown in figures 7.10 and 7.11.

subjects are shown in figures 7.10 and 7.11.

7.3.2 Subjects, Imaging and Data Processing

Whole-brain DWI datasets were acquired from 68 healthy volunteers (30.05 \pm 7.05 years, 10 Male) on a Siemens Trio 3T scanner with $1.71 \times 1.71 \text{mm}^2$ in-plane resolution, 2mm thick slices, six unweighted images and 64 diffusion weighted images ($b=1000 \text{s/mm}^2$) acquired with non-collinear diffusion sensitizing gradients.

DTI images for each subject were computed and deformably registered, using DTI-DROID [Yang et al.: 2008], to a DTI atlas [Wakana et al.: 2004]. Full brain tractography was performed following [O'Donnell and Westin: 2007], streamline tractography was performed by seeding in sub-voxel resolution by taking every voxel with linear anisotropy higher than .3, dividing it into $0.25 \times 0.25 \times 0.25 \text{mm}^3$ sub-voxels and seeding from each sub-voxel. In average, 10.000 fibre tracts were obtained for each subject. Then, the previously presented clustering algorithm was applied to every subject individually. In order to extract major white matter tracts on every subject individually we performed set of queries, shown in table 7.1, over the dendrogram obtained from the clustering of each subject using the tract querying algorithm we introduced. The whole process is shown in figure 7.8

7.4 RESULTS

We applied the clustering and tract-querying procedure to a 68 subject database. In order to provide qualitative assessment, we show the results of our clustering and tract querying algorithms for 2 different selected subjects in figure 7.9. In this figures, we note that tracts obtained by means of our clustering-querying proce-

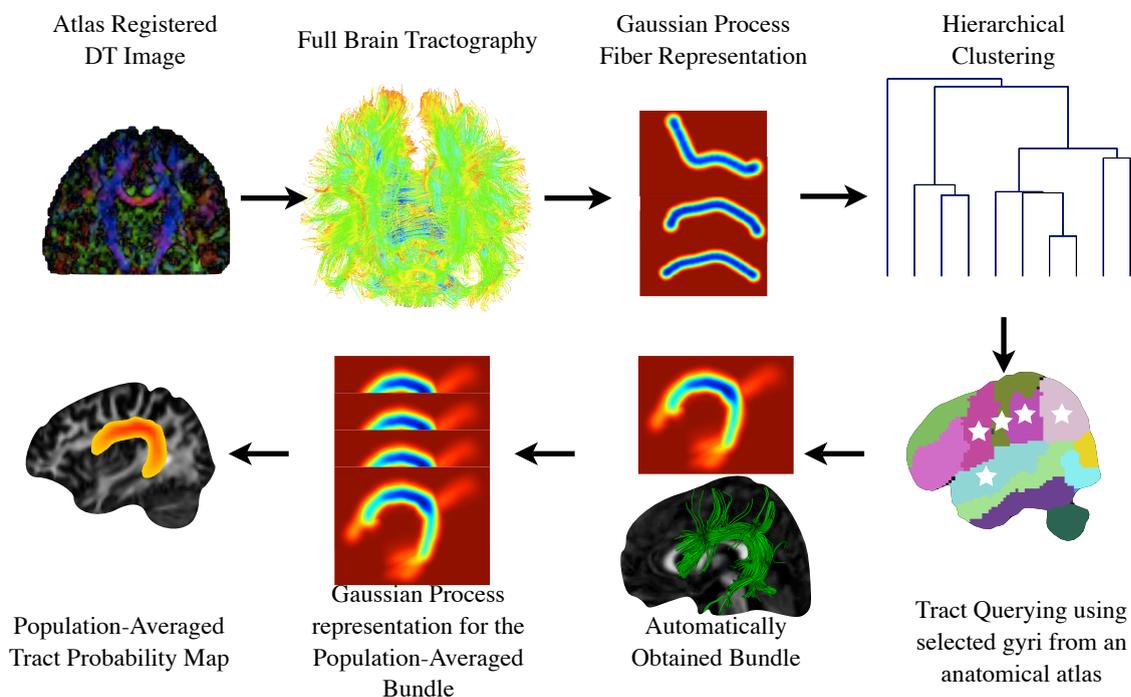


Figure 7.8: The procedure used to cluster white matter fibres into anatomical bundles and produce the *tract probability maps for each bundle*: We registered 68 Diffusion Tensor MR images using [Yang et al.: 2008]. We performed full brain tractography obtaining around 10.000 fibres per brain. We produced the Gaussian Process representation for each fibre as we describe in chapter 7. We identified anatomical bundles like the arcuate or the uncinate fasciculus by applying our clustering and tract querying algorithms to each subject individually as we describe in chapter 7. Finally we produced a population-averaged Gaussian Process for each identified bundle and the corresponding *tract probability map* with the methodology described in chapter 7

dures, are consistent with manually obtained tracts by experts in diffusion MRI images [Wakana et al.: 2004, Figures 3, 4, and 5] and macroscopical preparations [Lawes et al.: 2008, Figures 5iii and 6ii]. Furthermore, since our similarity measure handles partial overlap of fibres, we are able to correctly cluster fibres diverging from complex tracts like the arcuate fasciculus and the cingulum which innervate cortical and subcortical regions. This is observable in figure 7.9.

Then, we calculated the population-averaged *tract probability maps* for each bundle in order to evaluate cluster coherence across subjects. As the first step of this process we calculated the GP representation of the population-averaged bundle across the 68 subjects for each automatically extracted tract: For each queried white matter tract, the GP corresponding to the population-averaged bundle across subjects was calculated using equation 7.5. Then, we used equation 7.6 to calculate the *tract probability map* for each for each population-averaged bundle. We show the results of this step in figures 7.10 and 7.11, these maps are in atlas space, over fractional anisotropy images. Visual inspection of these *tract probability maps* shows that the population-averaged bundles are in agreement with probability maps obtained by manual selection of the bundles in Hua et al. [2008, Figure 3] and chemical staining

White matter tract to extract	Tract queries
frontal forceps	middle frontal orbital left gyrus, middle frontal orbital right gyrus
posterior forceps	medial occipital left gyrus, medial occipital right gyrus
fornix (left-section)	fornix left, medial temporal left gyrus
cortico spinal tract (post-central)	midbrain, post central gyrus
uncinate fasciculus	middle frontal orbital gyrus, medio temporal gyrus
arcuate fasciculus	post central gyrus, pre central gyrus, angular gyrus, supra marginal gyrus, superio temporal gyrus
cingulum bundle	cingulate gyrus
inferio fronto occipital tract	medial occipital gyrus, inferio frontal gyrus

Table 7.1: Queries applied to a set of 68 registered subjects in order to automatically extract major white matter fibre bundles using the algorithm presented in chapter 7. The results of this queries on 2 subjects are exhibited in figure 7.9 and *tract probability maps* for the *tract probability maps* corresponding to the population-averaged bundles for every subject are shown in figures 7.10 and 7.11

on post-mortem brains [Bürgel et al.: 2006].

Finally, in order to provide quantitative evaluation, for each bundle extracted from each subject we quantified its similarity with respect to the corresponding population-averaged bundle using our normalized similarity metric, equation 7.4. The result is plotted in figure 7.12. In this plot, the boxes span between the second and third quartiles of the similarity with the population-averaged bundle, the red bar is the median similarity. Moreover, the whiskers indicate the bundles whose similarity value with the population-averaged bundle is the smallest and largest within within 1.5 times the interquartile distance of the population-averaged similarity and the ‘+’ symbols indicate outliers. The high mean similarity with low dispersion on the cingulum, the cortico spinal tract and the inferior fronto-occipital fasciculus is consistent with histological studies [Bürgel et al.: 2006]. Additionally, the relatively lower and more disperse mean similarity on the arcuate and uncinate fasciculus is also consistent with previously cited histological studies.

7.5 DISCUSSION

Results showed that our clustering and tract querying method automatically differentiates white matter fibre bundles consistently across subjects. Furthermore, results demonstrate that we are able to identify white matter structures that agree with several works which manually perform white matter fibre bundle identification. Firstly, results of individual subjects are consistent with manually obtained tracts by experts [Wakana et al.: 2004] and macroscopical preparations [Lawes et al.: 2008]. Next, population-averaged *tract probability maps* match previously reported results obtained manually [Hua et al.: 2008] and through chemical staining [Bürgel et al.: 2006]. Finally, inter-subject tract variability measured through similarity

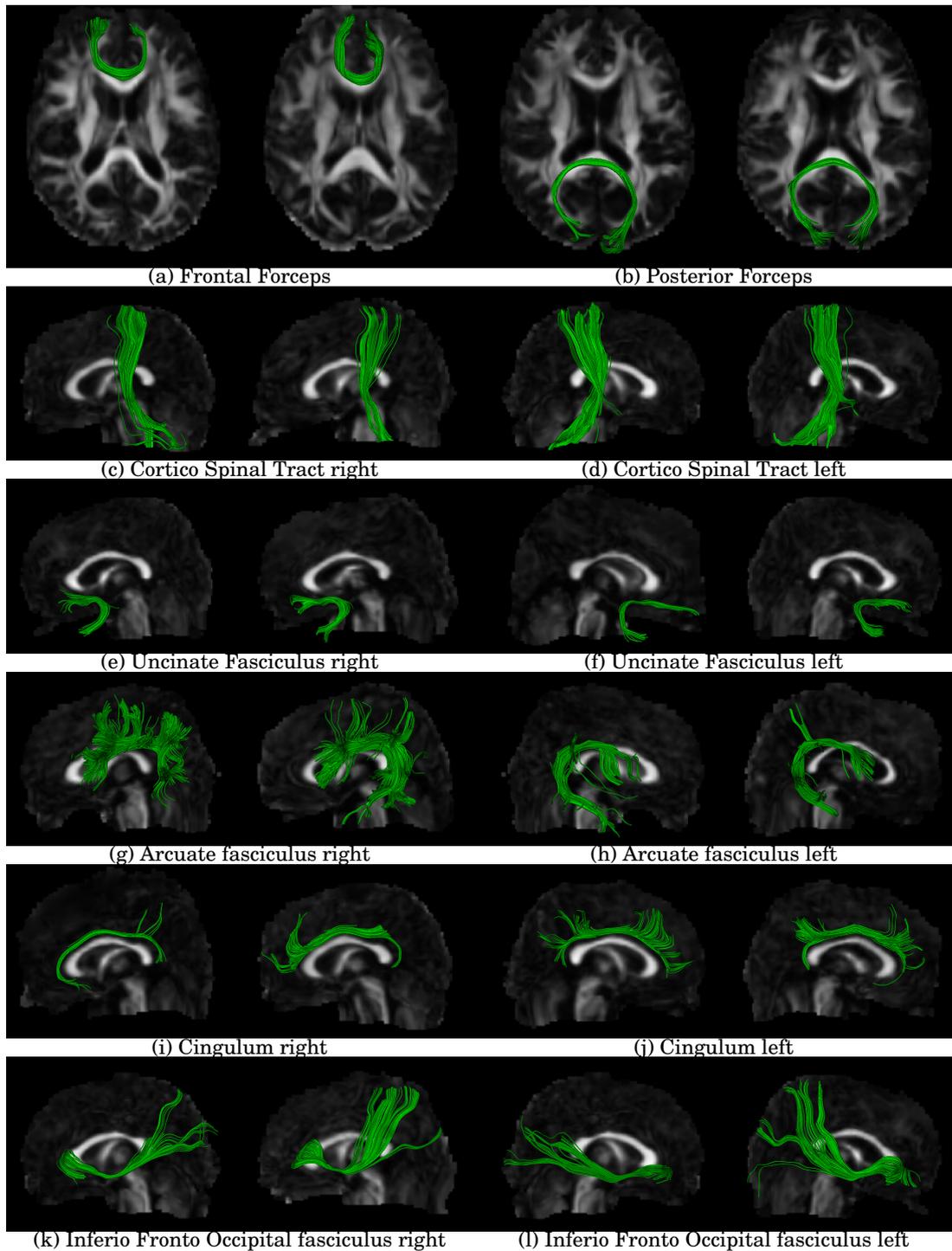


Figure 7.9: White matter fibre bundles automatically clustered from a database of 68 subjects using the clustering algorithm presented in chapter 7. The subjects were previously registered to a white matter fibre atlas and full brain tractography was performed by densely seeding on the whole cerebral white matter. Two subjects are shown.

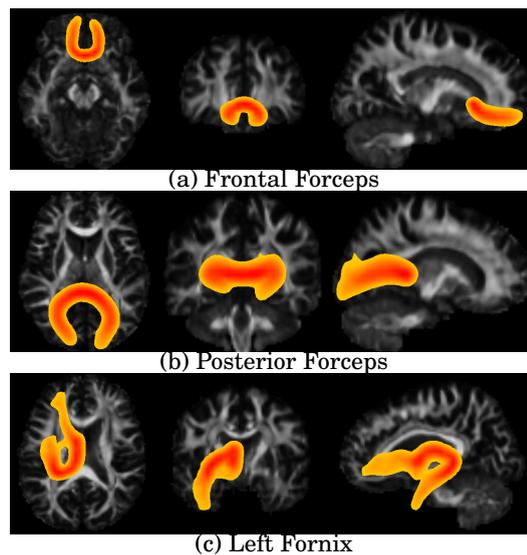


Figure 7.10: *Tract probability maps* for the averaged white matter fibre bundles over 68 subjects. The bundles have been automatically extracted from each subject individually using the clustering algorithm presented in chapter 7 and the queries presented in table 7.1. The subjects were previously registered to a white matter atlas and full brain tractography was performed by densely seeding on the whole cerebral white matter. Maximum intensity projection is used for the color intensity. Color code ranges from red, when the probability of the voxel belonging to the bundle is 1.0 to yellow, when the probability of the voxel belonging to the bundle is 0.2.

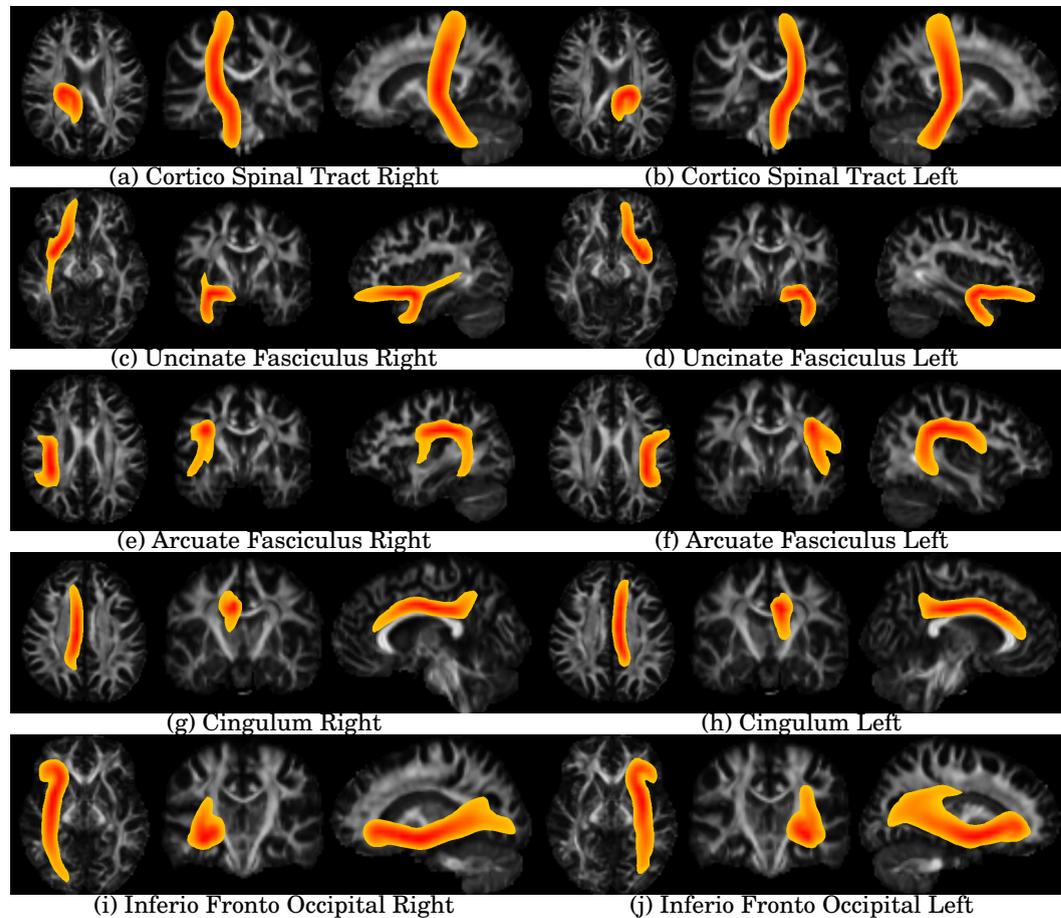


Figure 7.11: *Tract probability maps* for the averaged white matter fibre bundles over 68 subjects. The bundles have been automatically extracted from each subject individually using the clustering algorithm presented in chapter 7 and the queries presented in table 7.1. The subjects were previously registered to a white matter atlas and full brain tractography was performed by densely seeding on the whole cerebral white matter. Maximum intensity projection is used for the colour intensity. Color code ranges from red, when the probability of the voxel belonging to the bundle is 1.0 to yellow, when the probability of the voxel belonging to the bundle is 0.2.

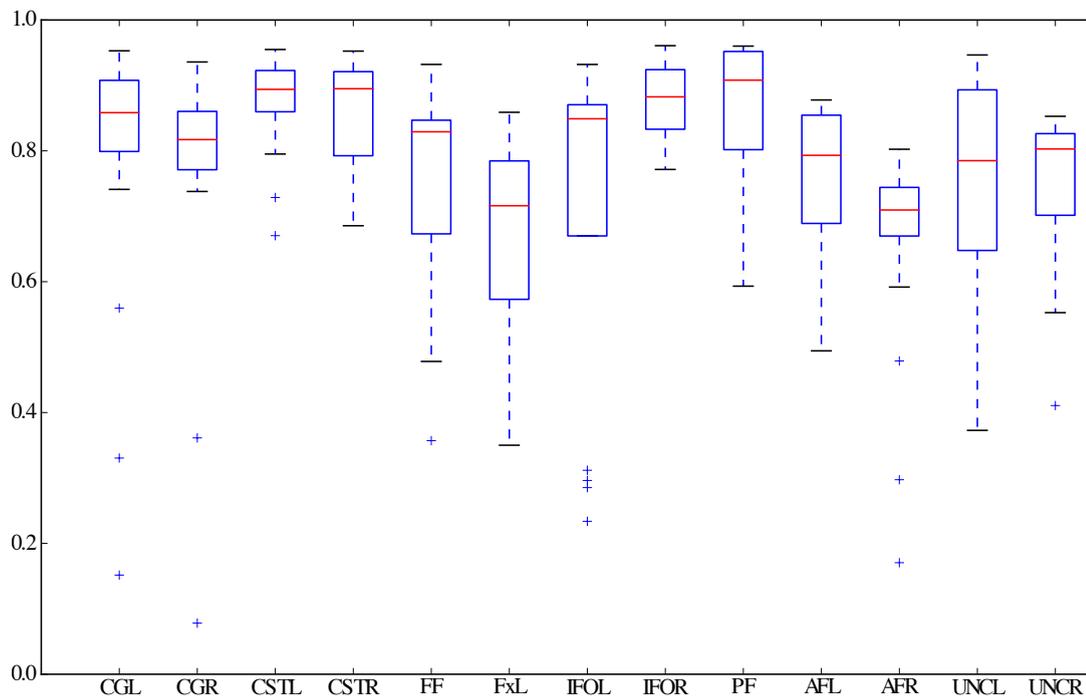


Figure 7.12: Quantitative assessment of bundle coherence among subjects. For each automatically extracted bundle, the similarity between the mean bundle across all subjects and the bundle extracted from each subject individually was calculated by means of the normalized inner product, equation 7.4. Similarity ranges from 1, bundles are identical, to 0, bundles are completely different. Boxes span between the second and third quartiles of the similarity with the mean bundle, the red bar is the median similarity. Moreover, the whiskers indicate the bundles whose similarity value with the mean bundle is the smallest and largest within 1.5 times the interquartile distance of the mean similarity and the '+' symbols indicate outliers.

with the population-averaged bundle is consistent with histological studies [Bürgel et al.: 2006]. Thus, thanks to our mathematical framework we are able to build an algorithm which performs automatic identification of white matter structures given a full-brain tractography.

In recent years tractography has become a popular means of performing white matter studies through diffusion MRI [Ciccarelli et al.: 2008]. Tractography results are visually appealing and recent developments produced means to perform analyses of diffusion-derived measures such as the fractional anisotropy. Still, statistical analysis on the fibres themselves has not been performed in sound ways due to the lack of an appropriate mathematical framework. Several examples of statistical analysis of white matter bundles including fibre clustering [O'Donnell and Westin: 2007, Wassermann and Deriche: 2008, Maddah et al.: 2008a] and tract probability maps [Hua et al.: 2008] have been reported. However, clustering approaches were either not suited for large-scale clinical studies or required a great deal of parameter tuning. With regards to *tract probability maps*, they were calculated by averaging binary masks obtained by manually extracted white matter fibres from tractography or by chemical staining.

This chapter proposes a mathematical framework that provides the necessary tools to perform automatic clustering and identification of white matter fibres and subsequent statistics on them. Automatic clustering of white matter fibres into bundles which produces consistent results between subjects is a fundamental tool for clinical studies. In this work we provide the tools to solve these issues. First, we develop a mathematical model for handling white matter bundles which includes spatial and diffusion tensor information and provides the grounds for their statistical analysis. Finally, we use it to develop a clustering algorithm and automatically obtain white matter structures which we then applied to 68 subjects.

7.5.1 Mathematical framework for white matter bundles

Our mathematical framework sets the foundation for statistical analysis of white matter fibre bundles including applications like clustering and tract-based quantification of scalar quantities amongst others. We provide three important operations for white matter bundle statistics: 1) the first operation to linearly combines white matter fibres into a bundle, see figure 7.2; 2) the second quantifies bundle similarity based on their overlap in space, see figure 7.3; and 3) the third calculates the probability that a point in space belongs to a bundle, called the *tract probability map* of the bundle.

Within our model, we represent white matter fibres and bundles as *blurred indicator functions*; Gaussian Processes are a parametrical representation of these functions. It is important to note that any GP representation of the *blurred indicator functions* will be able to implement the same operations as ours. These different GP representations are characterized by their covariance functions, which we present

in detail in chapter 7. Particularly, our choice of covariance functions contains our smoothness hypothesis about the fibres and enables us to perform efficient calculation of bundle similarity and *tract probability maps*. We chose a combination of two covariance functions to represent white matter bundles: one representing the smoothness of the bundles and a second changing blurring characteristics according to diffusion tensor information.

The smoothness covariance function allows us to infer the value of the *blurred indicator functions* from the sequences of points which represent each fibre. Three main characteristics led us to the choice of this covariance function. To begin with, it enforces the least wiggled representations of the fibres which has a smoothing effect on their trajectory. Secondly, it only enforces differentiability up to the second derivative, the minimum necessary to perform this smoothing. These characteristics lead to a covariance function equivalent of a compact support thin-plate spline function, that accurately models smooth spatial data [Wahba: 1990]. Other choices of smoothing covariance functions, like the widely used squared exponential function [Rasmussen and Williams: 2006], may lead to infinitely differentiable *blurred indicator functions*. It has been argued that covariance functions with this characteristic have undesirable geometric effects when representing implicit curves [Williams and Fitzgibbon: 2007] and are unrealistic for modelling many physical processes [Stein: 1999]. Finally, an important characteristic of our choice of covariance function is that it has compact support. This is not only more efficient computationally, but also leads to compact support *blurred indicator functions* which has an important effect on our similarity measure. Our operation quantifies the volume of the overlapping region between two fibres and can be normalized in order to provide a similarity index ranging from 1, when fibres overlap completely, to 0, when they are completely different. Compactness allows the similarity of two *blurred indicator function* which are sufficiently separated in space to be 0.

We combined the diffusion associated covariance function with that of smoothness in order to alter the blurring of the indicator function. Hence, the blurring is adapted according to diffusion tensor information without changing the compact support characteristic. Such covariance function is based on the diffusion tensor model, however in further work it would be simple to change it in order to make this model appropriate for more descriptive representations of the diffusion propagator [Tuch: 2004, Descoteaux et al.: 2007a]. This would be done by simply changing equation 7.8 to suit the chosen representation.

As a final word on the covariance functions, it is important to note that this model is flexible. In further work, a different analysis of white matter fibres and bundles could require a representation of the white matter bundles with a different set of hypotheses. In this case, a different set of covariance functions can be chosen or combined with the ones used in this work in order to retain most of the properties of our proposed framework while expressing different characteristics of the white

matter bundles.

7.5.2 Applications

We presented two applications of our Gaussian Process framework: clustering of white matter bundles and *tract probability maps*.

With regards to the clustering, we began by creating a dendrogram (see figure 7.6) from a full brain tractography. The dendrogram is a versatile tool for representing clustering results, given that it is a precomputed tree structure where each joint is a candidate cluster. Among other advantages, it allows for interactive exploration of the clustering results at different scales and is resilient to outliers. Moreover, clusters can be selected at different scales using different postprocessing methodologies. In particular, we take advantage of this feature using the *tract probability map* for each cluster and a volumetric atlas with a parcellation of the cerebral gyri. We combine both of these tools with the dendrogram in order to produce a *tract query* algorithm.

Tract probability maps were recently introduced as a tool to perform bundle-specific quantification of scalar quantities by Hua et al. [2008]. The cited work provides an example of this by analysing fractional anisotropy (FA) of a multiple sclerosis Patient in comparison with a normal population. Another example of this could be found by calculating the expected FA of a bundle, the weighted average of the FA using the values of the *tract probability map* as the weights. *Tract probability maps* have a wide range of applications. For instance, in this work we use them to relate the results of our clustering algorithm to anatomical information given by a white matter atlas enabling us to perform consistent identification of white matter structures among subjects. Although the clustering method is fully automated, it depends on the atlas parcellation and the method used to produce the tracts. We noticed two cases where non-anatomically coherent low probability regions (yellow) are present: the left section of the fornix, figure 7.10c, and the left uncinate fasciculus, figure 7.11c. In the case of the left side of the fornix, figure 7.10c, even though the high probability region (red) is consistent with the previously cited anatomical studies, there is a low probability region corresponding to the optical nerve that can be seen in this map. Careful analysis of the clustered fibres generating this region shows that this artefact comes from the tractography techniques used; these fibres go all the way from the posterior column of the fornix to the frontal sections of the optical nerve. This is not anatomically correct and is the source of the high variability exhibited by this tract in the quantitative analysis, figure 7.12. In the case of the left uncinate fasciculus, figure 7.11c, there is a high probability region (red) which is consistent with the previously cited anatomical studies. There is however, a low probability region corresponding to a section of the inferior-longitudinal fasciculus. This increases the variability of the clusters corresponding to the left uncinate fasciculus region, figure 7.12. Analysis of the clustered fibres constituting this bundle

and the query used to obtain the cluster corresponding to the uncinate fasciculus, table 7.1, shows that this artefact is produced by the white matter parcellation used to create the tracts. The absence of a parcel corresponding to the temporal pole in the atlas leads us to use the medial temporal gyrus for the corresponding query, shown in figure 7.7 right side in light green. This allows fibres fully contained within the medial temporal gyrus to be incorporated in the uncinate fasciculus as a result of the corresponding query. In the future, we will investigate the use of alternate atlases or expert-corrected ROIs to remedy this problem.

In further work, we plan to expand the applications of our framework. One of the most important potential applications is the use of tract probability maps as a probabilistic prior for streamline tractography. This would take advantage of the fact that Gaussian Processes are a generative model, thereby recovering tracts in the case of pathologies where the course of a white matter bundle might be interrupted. In turn, this would enable the comparison of a white matter structure predicted using control data to a pathological one.

7.6 CONCLUSION ---

We presented a mathematical framework to perform statistical operations on white matter fibres obtained from diffusion tensor MRI tractography. In doing so, we showed that this framework constitutes an inner product space which is appropriate for performing statistical operations among white matter bundles. We used this framework to build a clustering algorithm and a tract querying algorithm which allow automatic fibre bundle identification with the white matter queries as a sole parameter. Then, we applied these algorithms to 68 subjects registered to a publicly available atlas with a parcellation of the white matter gyri that we used as a priori anatomical information. Finally, we showed that the results of our clustering were consistent with studies carried by dissection macroscopical preparations and chemical staining. Thus, our framework has proved suitable to perform group studies of the cerebral white matter with minimal user interaction.

TRACTOGRAPHY-BASED SPINAL-CORD STRAIGHTENING

In this chapter we present an application of the Gaussian process framework for fibre bundles to spinal cord MRI images straightening. Spinal cord MRI (SC-MRI) is a challenging research field with numerous important clinical and basic research applications. Some of the SC-MRI applications strongly need to deal with a well straightened spinal cord either for appropriate methodological developments, for better visualization or diagnostic purposes. In this article, we develop an efficient and automatic method to straighten the spinal cord image and fibres. Diffusion Tensor MRI is first used to recover by tractography the bundles of fibres related to the spinal cord. The novel Gaussian process framework proposed in this thesis is then used to automatically recover in a robust way the most representative fibre which is used to interpolate and straighten the spinal cord image and fibres. Our method is successfully tested on real images of one cat with partial spinal cord injury and two healthy volunteers. This capability to reliably reconstruct straightened animal and human spinal cord opens new opportunities for SC-MRI applications.

8.1 INTRODUCTION

Spinal cord magnetic resonance imaging (MRI) has many applications in central nervous system diseases, such as multiple sclerosis and spinal cord injury. In clinical practice, radiological examinations aim at finding abnormalities through visual assessment. However, the curvature of the spinal cord in the antero-posterior (A-P) direction makes it difficult to obtain a full coronal or sagittal picture of the spinal cord which sometimes may hamper the diagnosis. Moreover, the curvatures of the spinal cord are subject-dependent and therefore group studies are hard to achieve at the spinal level. These observations highlight the need for a method to straighten the spinal cord so that individual images can be compared and eventually pooled.

A recent study has proposed a method to straighten the cord [[Stroman et al.: 2005](#)]. It is based on a manual delineation of the spinal cord, followed by a 3D reconstruction of the volume by interpolating each slice perpendicular to the curve. Al-

though very straightforward, this procedure requires manual delineation of the cord, which is time consuming and user-dependent. Moreover, only a planar delineation of the cord is achieved, providing straightening of the cord in only two dimensions of space, i.e., antero-posterior and rostro-caudal directions. If the subject is not well positioned or has scoliosis, curvature along the right-left (R-L) direction is likely to occur. In such case, the manual delineation of the cord becomes even more complex.

In this chapter we propose a fully-automatic method to straighten the spinal cord, based on diffusion tensor imaging (DTI) tractography and our novel Gaussian process framework. We used fibres of the spinal cord to generate a subject-dependant coordinate system. Hence, a key step in our algorithm is the robust selection of the most representative, or prototype, fibre from the set of white matter fibres traversing the spinal cord.

As a result of this work, the proposed method is robust and allows to straighten the cord in the A-P as well as in the R-L directions simultaneously.

8.2 MATERIAL AND METHODS

8.2.1 Theory

The first step to straighten the spinal cord using DTI tractography in our method is the selection of a prototype fibre, a representative fibre of the ensemble. This fibre is then used in order to produce a fibre-based coordinate system for the image.

Selecting the prototype fibre Although intuitively trivial, finding a prototype fibre describing a path along a large number of fibres is not an easy task. It should also be realized that fibres along a longitudinal tract may enter or leave the cord at different levels so that finding a representative tract or fiber bundle is not trivial. This is mainly due to the lack of a proper space to perform statistics among fibres. Recently, three methods were proposed to solve this problem: Maddah [Maddah et al.: 2007] chooses the longest fibre in the bundle, however this method is sensitive to long fibre outliers. To avoid this, Batchelor [Batchelor et al.: 2006] chooses the fibre with the greatest length weighted by local *fibre density* which requires calculating as the number of tractography trajectories passing through each voxel and to integrate this density along each fibre. Finally, O'Donnell [O'Donnell et al.: 2009] chooses the most representative of the fibres by performing spectral embedding of the fibres producing a point cloud. Then, the point closest to the mean of the cloud represents the prototype fibre. These techniques rely on point-to-point correspondences among fibres. Moreover, they require parameter tuning like the integration step over the fibres [Batchelor et al.: 2006] or scale and number of embedding dimensions [O'Donnell et al.: 2009].

To provide a method which does not rely on point-to-point correspondences, em-

bedding or parameter tuning, we build on the GP framework for fibre bundles presented in chapter 7. We represent each fibre \mathcal{F} as a parametric stochastic process $y(\cdot)$ whose realizations are blurred indicator functions on the image space. The value of these functions, $y(\mathbf{p})$, is 0 when there is no chance that the fibre traverses the point \mathbf{p} and 1 when the fibre surely traverses that point (figure 7.3). Then, if the most probable realization of $y(\cdot)$, $y^*(\cdot)$, is square integrable and has finite support, we define the similarity between two fibres, \mathcal{F} and \mathcal{F}' as the inner product

$$\langle \mathcal{F}, \mathcal{F}' \rangle := \int_{\mathbb{R}^3} y_{\mathcal{F}}^*(\mathbf{p}) y_{\mathcal{F}'}^*(\mathbf{p}) d\mathbf{p},$$

which roughly represents the volume of the overlapping between two blurred fibres. Moreover, we derive a fibre bundle norm from this inner product operation:

$$\|\mathcal{F}\|^2 := \langle \mathcal{F}, \mathcal{F} \rangle;$$

using the linear combination operations defined in chapter 7, we define a metric to measure dissimilarity between two fibre bundles

$$\text{dist}^2(\mathcal{F}, \mathcal{F}') := \|\mathcal{F} - \mathcal{F}'\|^2.$$

To select the prototype fibre from the ensemble we use the previously presented tools. We start by formulating the median fibre \mathcal{F}^* from an N fibre bundle \mathcal{B} in the presented fibre inner product space:

$$\mathcal{F}^* = \underset{\mathcal{F} \in \mathcal{B}}{\text{argmin}} \sum_{\mathcal{F}' \in \mathcal{B}} \text{dist}^2(\mathcal{F}, \mathcal{F}').$$

If we unroll the previous expression using the parallelogram theorem and the law of cosines

$$\mathcal{F}^* = \underset{\mathcal{F} \in \mathcal{B}}{\text{argmin}} \sum_{\mathcal{F}' \in \mathcal{B}} \|\mathcal{F}\|^2 + \|\mathcal{F}'\|^2 - 2\langle \mathcal{F}, \mathcal{F}' \rangle$$

it leads to

$$\mathcal{F}^* = \underset{\mathcal{F} \in \mathcal{B}}{\text{argmin}} N\|\mathcal{F}\|^2 + \sum_{\mathcal{F}' \in \mathcal{B}} \|\mathcal{F}'\|^2 - \sum_{\mathcal{F}' \in \mathcal{B}} 2\langle \mathcal{F}, \mathcal{F}' \rangle = \underset{\mathcal{F} \in \mathcal{B}}{\text{argmin}} N\|\mathcal{F}\|^2 - \sum_{\mathcal{F}' \in \mathcal{B}} 2\langle \mathcal{F}, \mathcal{F}' \rangle.$$

were the choice of the median fibre depends on its volume and overlapping with the rest of the bundle. From the previous expression we interpret the result of median operation geometrically: the resulting fibre is the one minimizing its volume, or length, and maximizing its overlapping with the rest of the fibres in the bundles. As what we are looking for is a fibre to parametrize the whole spinal cord, we drop

the volume constraint and we formulate the prototype fibre \mathcal{P} of the bundle as

$$\mathcal{P} = \operatorname{argmax}_{\mathcal{F}} \sum_{\mathcal{F}' \in \mathcal{B}} \langle \mathcal{F}, \mathcal{F}' \rangle \quad (8.1)$$

The expression to retrieve \mathcal{P} , equation 8.1, privileges the selection of a fibre from a dense pack ruling out fibres which do not have overlapping with the densest pack of the set. Hence, this selection procedure is robust to false positives in \mathcal{B} when they are small sets located outside the main structure in the same way that the median is robust to outliers.

8.2.2 Straightening the cord

Following acquisition of diffusion weighted (DW) data, tractography was performed using MedINRIA [?]. The tractography algorithm is based on a standard streamline algorithm [Mori et al.: 1999] and uses a tri-linear log-Euclidean interpolation method. The tractography procedure is fully automatic, since every voxel of the reference volume is a seed-point [?]. As a result, most fibres running along the spinal cord were reconstructed. This method is fairly robust due to the large presence of longitudinal tracts in the spinal cord.

During the next step, a robust algorithm is launched to retrieve the prototype fibre running along the spinal cord. This algorithm, described in subsection 8.2.1, has been designed to retrieve the prototype fibre among all fibres running along the spinal cord. This fibre is described as a 3D parametric function, allowing to perform arbitrary re-sampling on it. To compute this prototype fibre, an important concern was the robustness of the algorithm. It should not be sensitive to false positives, resulting from fibres located outside the spinal cord. This usually happens when structures external to the central nervous system are included in the field of view. Examples are the liver, the heart or the kidney, which are anisotropic structures. Another source of false positives are $N/2$ ghosting artefacts, which replicate the image along the phase-encoding direction. These observations highlight the need for a robust estimation of the prototype fibres using local priors (e.g., smooth curvature, packed geometry of fibres along the cord). These priors have been described in the previous section.

Following estimation of the prototype fibre, data are reconstructed by interpolating every slice perpendicular to this fibre. This is achieved by running an algorithm which starts at one extremity of the prototype fibre, and iterates by adding one arbitrary unit along the curved abscissa defined by the fibre. To reconstruct each slice, spline interpolation was used. The prototype fibre can be arbitrarily re-sampled. Therefore, the same non-linear transformation can subsequently be applied to any data that were acquired during the same scanning session. To achieve this, only information relative to voxel size and slice positioning are required. Then, a rigid-body

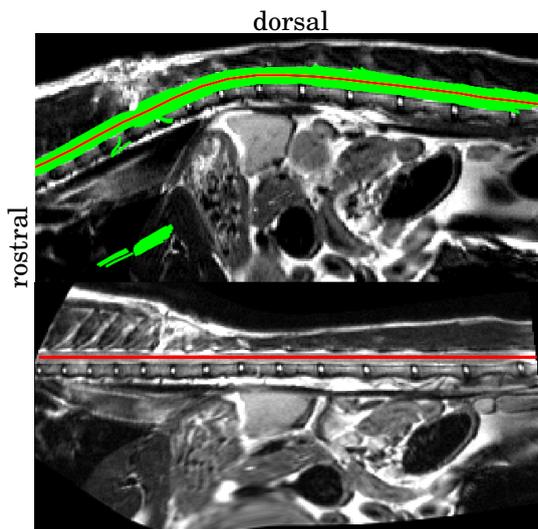


Figure 8.1: Results of our method to straighten the spinal cord. Top, sagittal slice of an anatomical image of a cat with the tractography results in green and the prototype fibre \mathcal{P} , selected using equation 8.1, in red. Bottom, straightening results.

transformation is applied to transform the coordinates of the prototype fibre from the DTI space into the anatomical space.

8.2.3 In vivo experiments

We validated the straightening method in one cat with partial spinal cord injury [Cohen-Adad et al.: 2008] and two healthy volunteers. The experiments were conducted in accordance with the Local Ethics Committee of the Université de Montréal (for animal studies) and the Université Paris 6 (for human studies). MRI acquisitions were carried out using a 3T MRI system (TRIO, 32 channel TIM system, Siemens) using phased array coils.

Images of the cat were acquired in the thoraco lumbar region (T6-L7). The anatomical scan consisted of a T2-weighted turbo spin echo (TSE) sequence (turbo factor of 13), $TR/TE = 7490/78ms$ at $0.55 \times 0.55 \times 1.1mm^3$ spatial resolution. DW data were acquired with a twice refocusing pulse single-shot spin echo echo planar imaging (EPI) sequence, iPAT=3 (to minimize susceptibility artifacts), sagittal orientation, 2 mm slice thickness, 128×128 matrix, $1.5 \times 1.5mm^2$ in-plane resolution, $TR/TE = 9500/109ms$, b-value = $1000s/mm^2$, 64 directions. We used respiratory gating to limit motion and susceptibility effects close to the lungs.

In humans, images were acquired in the cervico-thoracic region (C1-T4). The anatomical scan was a T2-weighted SPACE sequence with $TR/TE = 1500/120ms$ at $0.9 \times 0.9 \times 0.9mm^3$ spatial resolution. The DW sequence was similar to that one used in the cat, with $TR/TE = 3000/85ms$, $1.8 \times 1.8 \times 2mm^3$ spatial resolution, b-value= $1000s/mm^2$ and 60 directions, acquisition time <4 minutes. To demonstrate the robustness of the method, DW data were subsampled at 12 directions and processed the same way.

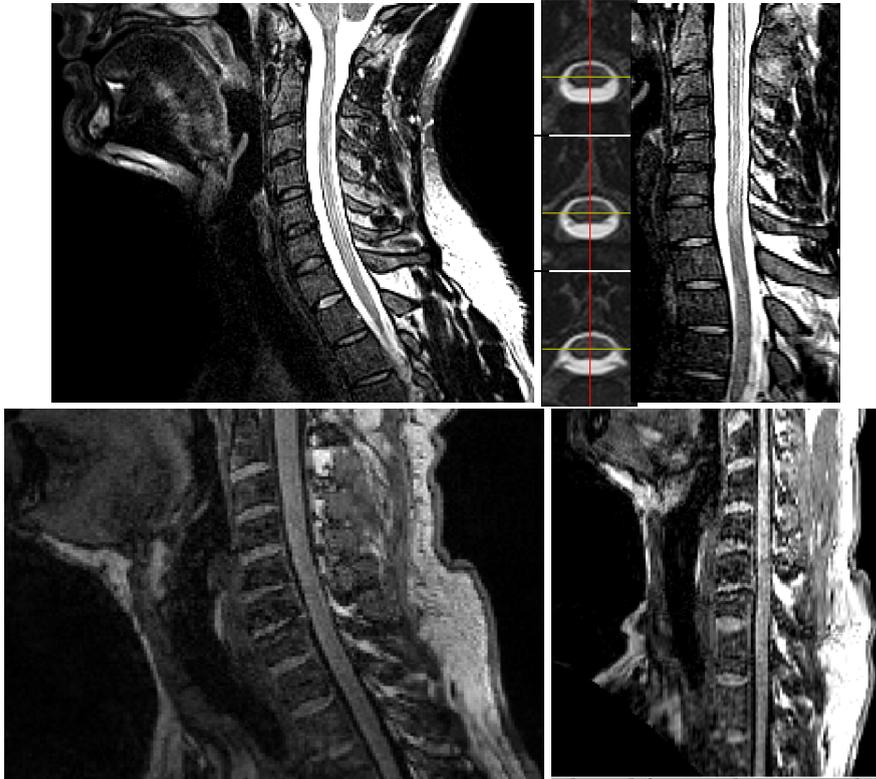


Figure 8.2: Results of our method to straighten the spinal cord for two human subjects. Left, sagittal slice of an anatomical image. Right, the straightened images are shown. Top center, three axial slices of the straightened image of the top subject shown to exhibit the three-dimensional alignment.

8.2.4 Data processing

Images were first averaged respectively to their diffusion directions and were two times interpolated. Series acquired without parallel imaging were corrected for susceptibility-induced distortions using the method described in [?].

8.3 RESULTS

One motivation of this work was to show that adding a short DW sequence to the MRI protocol, a fully automated method to straighten the spinal cord is feasible. This study therefore provides results of cord straightening in a cat and two human subjects. Figure 8.1 (top) shows a sagittal slice of the T2 image of the cat with the fibres obtained by tractography in green and the prototype fibre retrieved in red. The bottom row shows the result of the straightening procedure as described in subsection 8.2.2. Figure 8.2 shows the result for two human subjects: On the left, sagittal slices of T2 images for two human subjects. Their respective straightened results are shown on the right. In the case of the top subject, three axial slices of the straightened image are shown in the middle column to demonstrate the three-dimensional alignment.

8.4 DISCUSSION AND CONCLUSIONS

We presented a fully automatic method to straighten the spinal cord, allowing better visualisation of the structure and enabling to conduct group analysis. The method uses DTI tractography to delineate the spinal cord and an advanced algorithm to robustly find the central line along the spinal cord. The method has been tested in humans and cats at 3T, but can be applied to other mammals.

Robustness Here we demonstrated the feasibility of straightening the spinal cord with 60 directions DW acquisitions acquired in 4 minutes. To study the robustness of the algorithm with lower sampled DTI data, we sub-sampled the human DW dataset into 12 directions and applied the straightening method. Results showed a very satisfactory straightening of spinal the cord for both subjects (data not shown). A DW acquisition in twelve directions takes less than a minute and therefore could easily be plugged into an imaging protocol.

STATISTICAL ANALYSES OF THE WHITE MATTER IN SCHIZOPHRENIA

9.1 INTRODUCTION

In this chapter we illustrate the applicability of the Gaussian process framework introduced in chapter 7 to perform group analysis on white matter bundles. In the cited chapter, we developed techniques to automatically identify white matter structures for individual subjects and obtain population-averaged tract probability maps. The aim is to study group differences in white matter between healthy controls and patients with schizophrenia. This is achieved by using two of the most popular tract-based analyses in current literature.

In looking for group differences between brains of schizophrenia patients and controls, we perform two types of statistical analyses on a selected set of scalar quantities measuring different characteristics of the diffusivity profile within a voxel . These measures were chosen due to their biological interpretation and use in diffusion tensor studies of white matter differences in pathologies [Ciccarelli et al.: 2008, Kubicki et al.: 2007].

In the two types of statistical analyses, the first is inspired by the works of Paganini et al. [2005] and Lin et al. [2007]. This analysis combines a scalar diffusivity measure like the FA and the tract probability maps such as the ones we computed in chapter 7 to obtain a single-scalar diffusivity measure for each tract. The second, inspired by the popular TBSS methodology [Smith et al.: 2006], uses the tract probability maps to produce a skeleton¹ of the tract. This provides a tool to search for differences in specific sections of a white matter structure. The final step of both techniques relies on hypothesis tests in order to find differences between the controls and patients.

1. In this work we follow the same definition of skeleton as in Smith et al. [2006] where it can be shaped as a curve or a sheet.

9.2 STATISTICAL ANALYSIS ALGORITHMS

In this section we describe the two statistical analysis algorithms presented in chapter 9. Both algorithms are based on the tract probability map (TPM) function (equation 7.6) which is the probability that a voxel \mathbf{p} belongs to the bundle \mathcal{B} :

$$\text{TPM}(\mathbf{p}) = \mathbb{P}\{\mathbf{p} \in \mathcal{B}\} = \frac{c}{\sqrt{\pi(h^2 + \sigma_{\mathcal{B}}^2(\mathbf{p}))}} \quad (9.1)$$

where c is a constant so the TPM integrates to 1, h is a bandwidth parameter and $\sigma_{\mathcal{B}}^2(\cdot)$ is the variance of the Gaussian process representing the fibre bundle \mathcal{B} at \mathbf{p} .

9.2.1 Whole Tract Analysis

We analysed changes in scalar diffusivity measures, such as the FA, between controls and patients on the tracts as a whole. For this, we calculated the expectation of these quantities, according to the TPMs, for each structure in template space. An example of this is the expected axial diffusivity on the arcuate fasciculus (AF), which was calculated as

$$E_{\lambda_{\parallel}}^{AF} = \int \text{TPM}^{AF}(\mathbf{p}) \lambda_{\parallel}(\mathbf{p}) d\mathbf{p}.$$

In this equation, $\text{TPM}^{AF}(\mathbf{p})$ is the probability that the voxel \mathbf{p} belongs to the AF and $\lambda_{\parallel}(\mathbf{p})$ is the axial diffusivity at \mathbf{p} .

Such an expected value was computed for each of the scalar quantities for WM structures of choice. Each white matter structure was then analysed using two-tailed statistical tests on the data: 1) the t-test which assumes normality and 2) the Mann-Whitney which is non-parametric [Mann and Whitney: 1947]. The Mann-Whitney test has been extensively used in the community as a replacement of the t-test when normality can not be assumed [Pagani et al.: 2005, Lin et al.: 2007, Ciccarelli et al.: 2008].

9.2.2 Tract-Skeleton Analysis

We develop this statistical analysis in order to find differences in precise locations of the tracts. The algorithm to analyse a particular fibre bundle starts by generating its TPM as in chapter 7. By construction, the $\text{TPM}(\cdot)$ function has a ridge (or valley) of high probability at the centre of the tract that decays towards the exterior. Consequently, the main idea is to find a representation of the tract as a sheet or line and to project diffusivity information on that representation. For this, we adapt the thinning algorithm of Smith et al. [2006]. Then, we project a scalar diffusivity measure (e.g. the FA) on the skeleton and perform group analysis on it. This process is shown in figure 9.3.

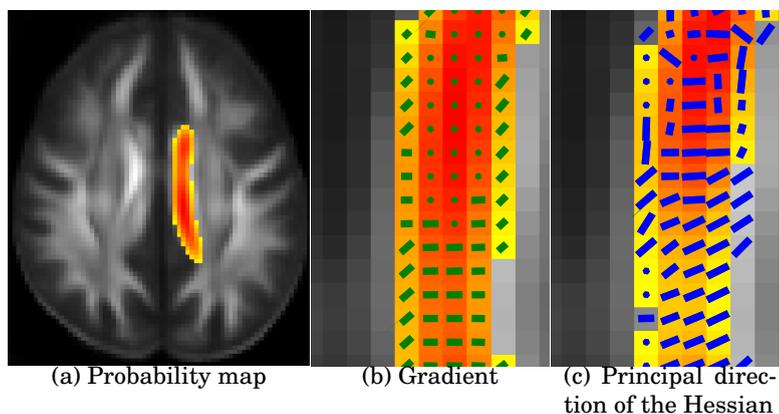


Figure 9.1: Finding the direction perpendicular to the tract probability map skeleton: When the voxel is far enough from the central ridge or valley, the gradient direction (b) is perpendicular. However, when it is close enough, the gradient norm is close to zero as shown by the circles in (b). In these cases, the principal direction of the Hessian indicates the perpendicular direction (c).

Skeleton Calculation

The first step in the algorithm is to obtain the skeleton. For this, we start by identifying the voxels containing such skeleton. We start by computing the centre of gravity (CofG) of a sphere of radius r around the voxel \mathbf{p} , $S(\mathbf{p}, r) \in \mathbb{R}^3$,

$$\text{CofG}(\mathbf{p}; r) = \int_{\mathbf{q} \in S(\mathbf{p}, r)} \mathbf{q} \text{TPM}(\mathbf{q}) d\mathbf{q}. \quad (9.2)$$

If the centre of gravity is sufficiently close to \mathbf{p} we assume that \mathbf{p} is very close to the ridge or valley which we will call skeleton from now on. Finally, we extract the skeleton. We select voxels which are very close to the skeleton as previously defined. Then, we analyse these voxels in a direction perpendicular to the skeleton. Those being a local maximum along this direction are assumed to be on it and extracted.

As we have seen previously, to extract the skeleton we need to calculate the direction that is perpendicular to it. If the voxels are sufficiently far away from it, this direction is provided by the gradient of the TPM. This is due to the fact that the probability of a voxel belonging to the underlying tract of the TPM decreases when we move away from the skeleton. However, if we are on the skeleton, meaning on a ridge or a valley, the gradient will have a 0 norm and no defined orientation. In these cases it is reasonable to assume that the gradient changes more rapidly in the direction perpendicular to the skeleton [Lindeberg: 1998]. Thus, we take the eigenvector corresponding to the largest eigenvalue of the Hessian of the TPM, as the direction perpendicular to the skeleton. This is illustrated in figure 9.1. Then, the direction perpendicular to the skeleton at \mathbf{p} is given by

$$\text{skel}_{\perp}(\mathbf{p}) := \begin{cases} \nabla \text{TPM}(\mathbf{p}) / \|\nabla \text{TPM}(\mathbf{p})\| & \text{if } \|\text{CofG}(\mathbf{p}; r) - \mathbf{p}\| > \epsilon \\ \mathbf{e}_1(\nabla^2 \text{TPM}(\mathbf{p})) & \text{if } \|\text{CofG}(\mathbf{p}; r) - \mathbf{p}\| \leq \epsilon \end{cases}. \quad (9.3)$$

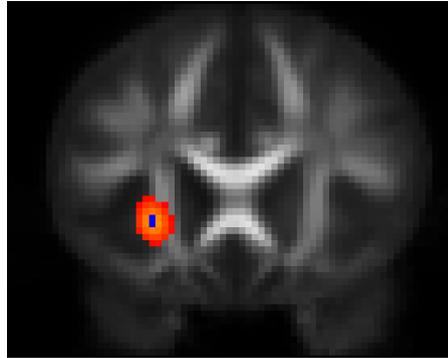


Figure 9.2: Voxels projected to a point in the skeleton: We show a point on the inferior fronto occipital tract skeleton in blue. The voxels in the image which are closest to it are colored in a red-yellow scheme where red is the lowest probability and yellow the largest. The values of these voxels are projected to the skeleton and averaged according to their probability of being in the bundle.

In this equation ϵ is taken to be a tenth of the voxel width and $\mathbf{e}_1(M)$ is the eigenvector corresponding to the largest eigenvalue of M , and ∇ and ∇^2 are the gradient and the Hessian operators. Moreover, due to the properties of our GP framework, the gradient and Hessian of the TPM (equation 9.1) can be calculated analytically.

Finally, it is important to emphasize that this is just one technique to calculate the skeleton. The statistical analysis algorithm is not dependent on it, is general and can be applied to any skeleton that is constructed. Most of the ridge and valley calculation algorithms are based on the gradient and Hessian of the analysed function [Lindeberg: 1998, Kindlmann et al.: 2009], hence it is possible to reformulate them in terms of the GP framework for white matter bundles.

Voxel Projections and Statistical Analysis

Having created the skeleton for a tract, we project all the voxels within the thresholded TPM to their closest point on the skeleton. This is achieved by starting at every voxel and following the direction perpendicular to the skeleton at that voxel, calculated in equation 9.3, until the skeleton is reached. Then, we calculate the expected scalar diffusivity measure at every point of the skeleton. In figure 9.2, we show a point on the inferior fronto occipital tract (IFO) skeleton in green and the voxels in the image which are closest to it. The values of these voxels are projected to the skeleton and averaged according to their probability of being in the bundle. This procedure for the IFO and the fractional anisotropy (FA), is as follows: for each point in the skeleton of IFO, \mathbf{s} , we calculated the set of voxels $\mathcal{N}(\mathbf{s})$ within the TPM having \mathbf{s} as the closest on the skeleton. This is done by following the direction perpendicular to the skeleton calculated in the previous section. Then, the expected value of FA on each skeleton voxel \mathbf{s} , is the expectation of the value of the FA on $\mathcal{N}(\mathbf{s})$ projected to \mathbf{s} :

$$E_{\text{FA}}^{\text{IFO}}(\mathbf{s}) = \frac{\int_{\mathcal{N}(\mathbf{s})} \text{TPM}^{\text{IFO}}(\mathbf{p}) \text{FA}(\mathbf{p}) d\mathbf{p}}{\int_{\mathcal{N}(\mathbf{s})} \text{TPM}^{\text{IFO}}(\mathbf{p}) d\mathbf{p}} \quad (9.4)$$

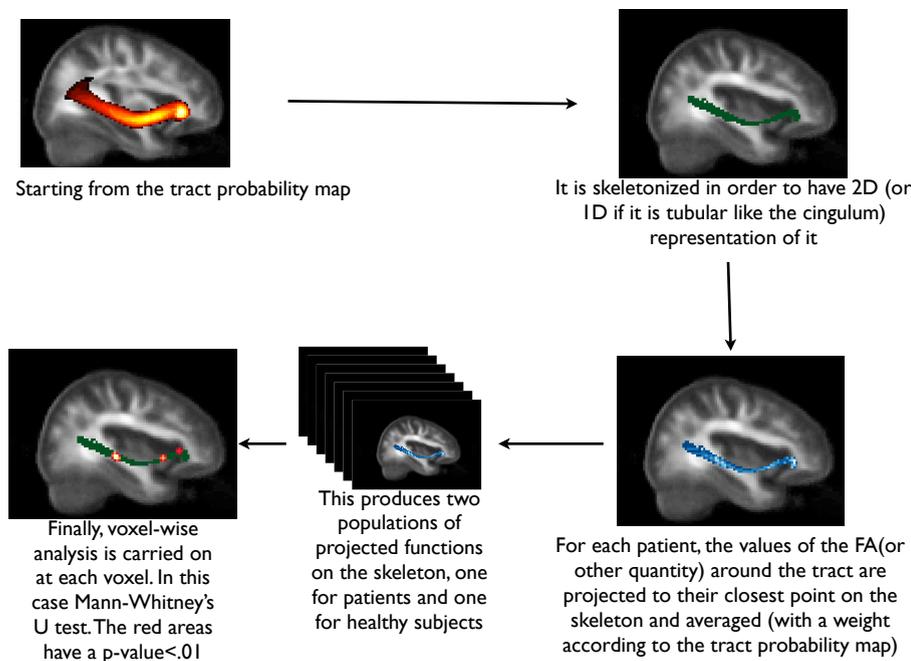


Figure 9.3: Statistical procedure to find differences along the tract: We start from the Tract Probability Map. The first step is to skeletonize it in order to reduce the dimensionality of the tract representation. Next, for each voxel in the skeleton, the weighted average of the FA (or other scalar diffusivity measure) among the voxels which are closer to that point is calculated. This leads us to a different projected function on the skeleton for each subject. Finally, voxel-based analysis is performed on the skeleton voxels.

Finally, for each voxel in the skeleton, we perform voxel-based analysis. In this case we use the Mann-Whitney U test to find voxels where the distribution of the fractional anisotropy or the axial or radial diffusivities among patients and controls is different, and we report trends indicated by the analysed t-score.

9.3 DATA

Whole-brain DWI datasets were acquired from 58 volunteers (35.60 ± 10.48 years, 28 Male, 24 healthy, 34 diagnosed with schizophrenia) on a Siemens Trio 3T scanner with $1.71 \times 1.71 \text{mm}^2$ in-plane resolution, 2mm thick slices, six unweighted images and 64 diffusion weighted images ($b=1000 \text{s/mm}^2$) acquired with non-collinear diffusion sensitizing gradients.

9.3.1 Data preparation

DTI images for each subject were computed and deformably registered, using DTI-DROID [Yang et al.: 2008], to a DTI atlas [Wakana et al.: 2004]. Full brain tractography was performed following [O'Donnell and Westin: 2007], streamline tractography was performed by seeding in sub-voxel resolution by taking every voxel with linear anisotropy higher than .3, dividing it into $0.25 \times 0.25 \times 0.25 \text{mm}^3$ sub-voxels and seeding from each sub-voxel. In average, 10.000 fiber tracts were obtained for each subject.

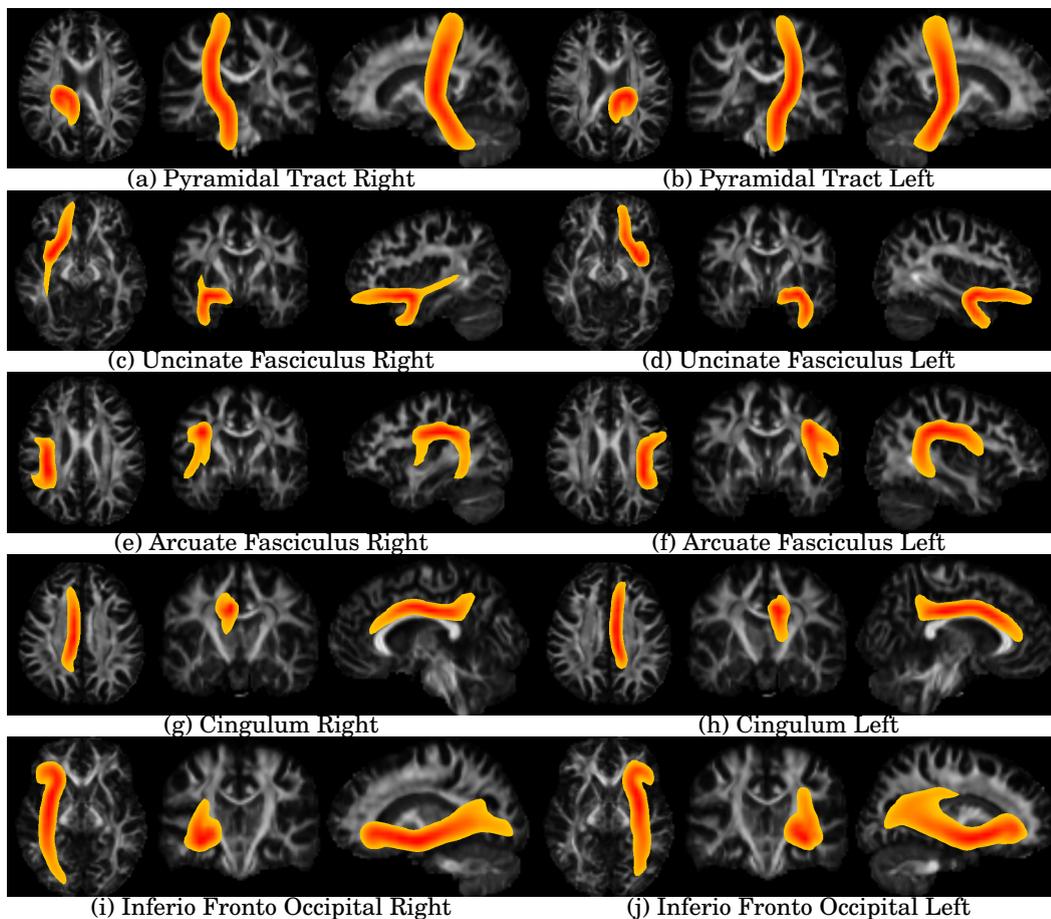


Figure 9.4: *Tract probability maps* for the averaged white matter fiber tracts over 21 subjects. The tracts have been automatically extracted using the techniques developed in chapter 7. Maximum intensity projection is used for the color intensity. Color code ranges from red, when the probability of the voxel belonging to the tract is 1. to yellow when it is 0.2.

9.3.2 White Matter Structure Identification

In order to identify white matter structures, the clustering algorithm presented in chapter 7 was applied to every subject individually. Population-averaged tract probability maps for major tracts were obtained for each extracted white matter structure: on both hemispheres the arcuate, inferio fronto occipital, uncinate and cingulum fasciculi along with the fornix; also, the frontal and posterior forceps were extracted. The obtained TPMs are shown in figure 9.4. The TPMs were analysed at three different scales: they were thresholded and re-normalized in order to include the 40%, 50% and 80% most probable area of each white matter structure respectively. This is illustrated in figure 9.5.

9.3.3 Statistical analysis of tracts

To characterize diffusion properties of the white matter structures, four scalar quantities derived from the diffusion tensors were used: Fractional Anisotropy (FA), axial diffusivity ($\lambda_{\parallel} = \lambda_1$) and radial diffusivity ($\lambda_{\perp} = \frac{1}{2}(\lambda_2 + \lambda_3)$). The FA measure was used due to its biological interpretation as a measure of tract integrity and its popularity in the literature. We chose the other two measures for two reasons, firstly their response to differences on the diffusivity along the eigenvalues of the diffusion tensor is linear (no clue what this means). Secondly, there are several reports relating differences on axial and radial diffusivity with differences in myelination and axonal calibre [Song et al.: 2002, 2005, Barazany et al.: 2009, Beaulieu: 2002, Concha et al.: 2006].

9.4 RESULTS

We performed the “whole tract analysis”, introduced in subsection 9.2.1, at different thresholds as illustrated in figure 9.5. In doing this using the Mann-Whitney U test, we detected trends of significant difference ($p\text{-value} < 0.01$) : an increase of λ_{\parallel} and λ_{\perp} in the left fornix and uncinate fasciculus, the frontal forceps, and the right and left pyramidal tracts; increased FA in the left inferior fronto occipital fasciculus; and a decrease of the FA on the left arcuate fasciculus and fornix.²

We also performed a “tract-skeleton analysis”, introduced in subsection 9.2.2. Results of this analysis are shown in figure 9.6. In this figure we show tracts where the number of voxels with a p -value under 0.001 is above the 10% of the number of voxels in the whole skeleton. For these, we report the trend indicated by the t -score in order to characterize the detected change in diffusivity properties: in the left fornix there are regions with an increase of λ_{\parallel} and λ_{\perp} ; the right arcuate fasciculus shows regions with a decrease of FA; the right inferior fronto occipital fasciculus shows areas of decreasing λ_{\parallel} ; in the left inferior fronto occipital fasciculus there are areas of increased axial and λ_{\perp} ; the frontal forceps shows areas of decreased λ_{\perp} ; the posterior forceps has areas of increased FA and λ_{\perp} ; and the left pyramidal tract exhibits an increase of λ_{\parallel} and λ_{\perp} .

9.5 DISCUSSION

Our results are in agreement with current literature. The predominance of differences in tracts within the left frontal and temporal lobes has been reported by a recent meta-analysis of DTI studies in schizophrenia [Ellison-Wright and Bullmore: 2009] and a detailed survey [Kubicki and Shenton: 2009]. Particularly, the decrease of the FA on the left arcuate fasciculus, shown by both of our analyses was reported by Burns et al. [2003]. Regarding our studies analysing FA changes in tract skele-

². Due to data disclosure conditions on the used database, we are not authorized to show quantitative results

Tract Name	Differences		Trend of the difference	Studies Agreeing
	Left	Right		
Inferior Fronto Occipital	✓	✓	$\lambda_{\parallel} \uparrow$ left only: $\lambda_{\perp} \uparrow$ FA \uparrow	Ashtari et al. [2007]
Arcuate Fasciculus	✓	×:	FA \downarrow	Burns et al. [2003]
Cingulum	×	×		
Fornix	✓	×	$\lambda_{\parallel} \uparrow$ $\lambda_{\perp} \uparrow$ FA \downarrow	Fitzsimmons et al. [2009]
Uncinate	✓	×	$\lambda_{\parallel} \uparrow$ $\lambda_{\perp} \uparrow$	
Pyramidal	✓	✓	$\lambda_{\parallel} \uparrow$ $\lambda_{\perp} \uparrow$	
Commisural Tracts				
Frontal Forceps		✓	$\lambda_{\perp} \downarrow$	
Posterior Forceps		✓	FA \uparrow $\lambda_{\perp} \uparrow$	

Table 9.1: Summary of results between controls and schizophrenics. Studies agreeing with our results are cited. The predominance of the differences in the frontal and temporal lobes of the left hemisphere is in agreement with a recent meta-analysis of DTI studies in schizophrenia [[Ellison-Wright and Bullmore: 2009](#)] and a detailed survey. [[Kubicki and Shenton: 2009](#)]

tons, we found an increased FA in left IFO which was previously shown by [[Ashtari et al.: 2007](#)] and decreased FA in the fornix is consistent with [Fitzsimmons et al. \[2009\]](#). This agreements show that our statistical analyses are suited for white matter studies encouraging us to carry on with a more comprehensive study on larger databases.

Regarding to the analyses of parallel (λ_{\parallel}) and perpendicular (λ_{\perp}) diffusivities, even if recent work shows that these measures are more easy to correlate to biological changes [[Beaulieu: 2009](#)], there are no schizophrenia studies using DMRI considering them [[Kubicki and Shenton: 2009](#), [Ellison-Wright and Bullmore: 2009](#)]. However, our results agree with [Ellison-Wright and Bullmore \[2009\]](#) in that the differences found are predominantly in the temporal and frontal lobes of the left hemisphere. In future work, we expect that an in depth study of the differences in these quantities will shed some light into the biological characteristics of white matter changes in schizophrenia. The results and comparison with present studies are summarized in table 9.1.

Finally, none of the methods presented is bounded to a single scalar diffusivity measure, they are even generalizable to full-tensor studies as done, for instance, by [Yushkevich et al. \[2008\]](#). Moreover, it is possible to correlate these changes with age and cognitive benchmarks in order to better analyse the relation between white matter changes and the different disorders provoked by schizophrenia.

9.6 SUMMARY

In this chapter we have used our novel GP framework for white matter fibre bundles in order to perform statistical analysis. For this, we have formulated statistical protocols widely used in the literature in terms of our GP framework. We have applied the statistical analysis protocols in a small database and showed that we found

trends of significant group differences. In turn, working on a larger database, our proposed GP-based analysis will be able to shed some light into group differences characterizing pathologies.

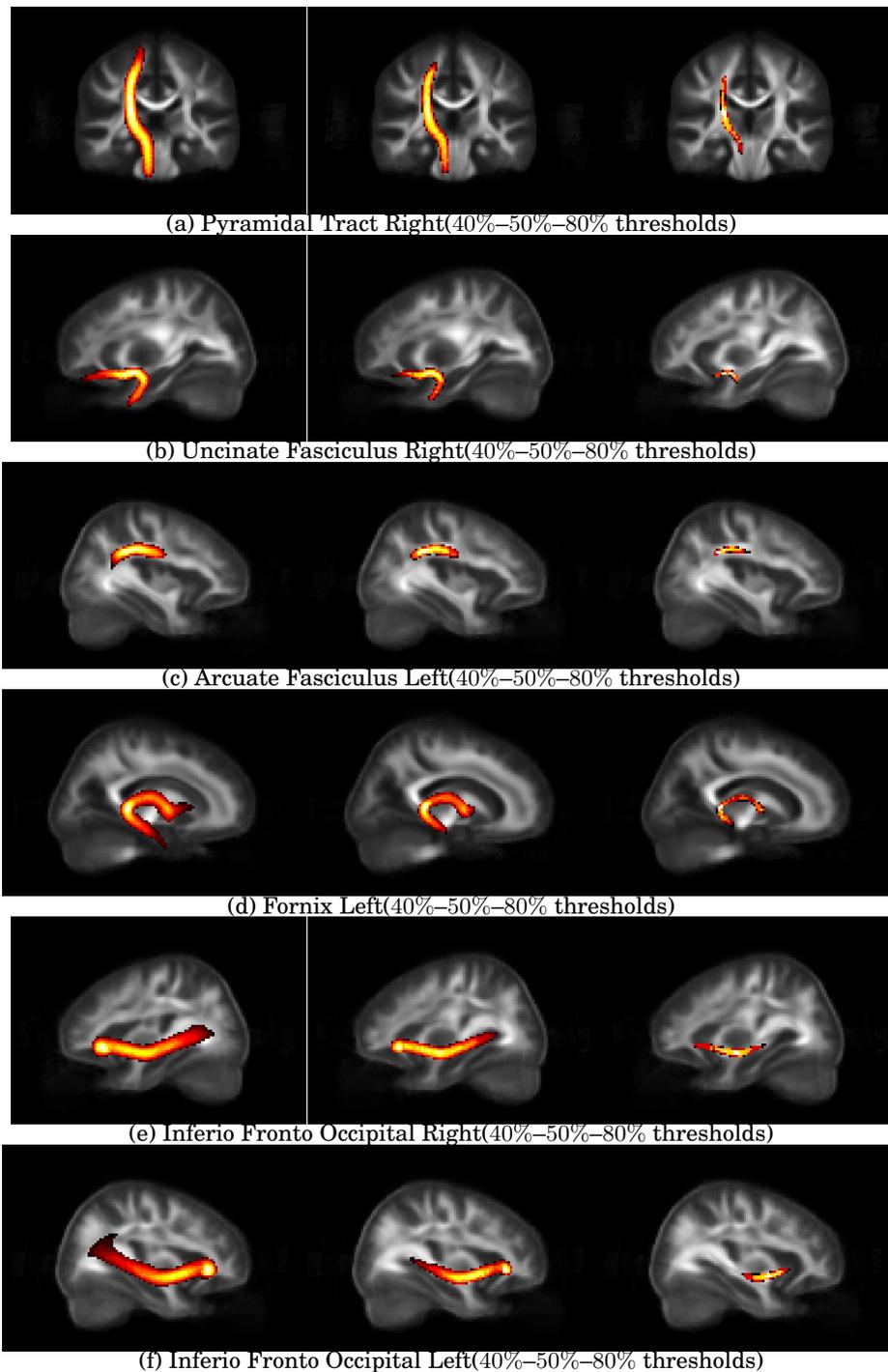


Figure 9.5: *Thresholded Tract probability maps* for white matter fiber tracts averaged over 21 subjects. The previously obtained tract probability maps are thresholded with respect to the maximal value to better localize the analysis. The thresholds on the left, center and right panel are 40%, 55% and 80% respectively. Maximum intensity projection is used for the color intensity. Color code ranges from red, when the probability of the voxel belonging to the tract is 1.0 to yellow, when the probability of the voxel belonging to the tract is 0.2.

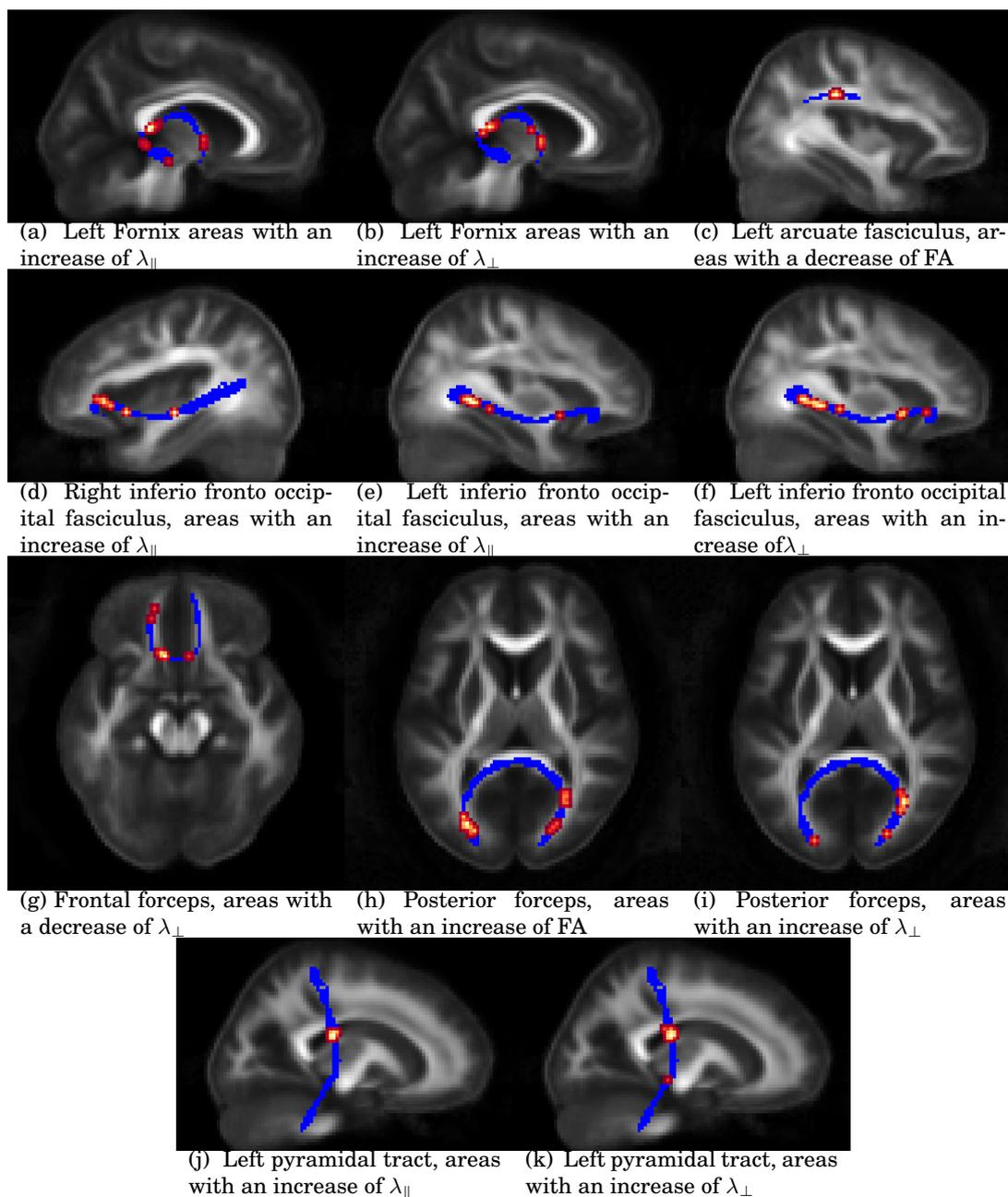


Figure 9.6: Differences in tract portions: Using the statistical procedure described in subsection 9.2.2, we show tracts where some sections had differences between controls and patients. Only tracts which demonstrate a trend towards a significant variation of the diffusivity quantities are shown. For these, we highlight the areas which show a variation and report the type of change.

Acknowledgements

The dataset used in chapters 7 and 9 was collected as part of NIH grants #MH-60722, MH49142, MH65578 and MH65578 by the imaging team of the Brain Behavior Lab, Neuropsychiatry section at UPenn. I thank Drs. R. C. Gur, J. Loughhead and M. Elliot for providing the scans. The processing was done at the Section of Biomedical Image Analysis, Department of Radiology, UPenn as part of NIH grant #MH079938. I thank Dr. R. Verma, Mr. S. Kanterakis and Mr. L. Bloy for the pre-processing.

The acquisition and preparation of the datasets used in chapter 8 was supported by the Canada Research Chair on the Spinal Cord provided by the Canadian Institute of Health Research (CIHR) to S. Rossignol, by the Multidisciplinary Team on Locomotor Rehabilitation (Regenerative Medicine and Nanomedicine, CIHR), by the association française contre les myopathies (AFM) and the institut pour la Recherche sur la moelle épinière et l'encéphale (IRME). I thank J. Cohen-Adad, H. Benali and S. Rossignol for their collaboration on this project and H. Leblond, H. Delivet-Mongrain and J. Provencher for the cat preparation, C. Hurst for assistance with MRI acquisition.

Conclusion

In this thesis, we were interested in proposing methods to analyse the human brain's white matter structure using diffusion magnetic resonance imaging (DMRI). More precisely, in performing automated *in vivo* dissection of the white matter and developing statistical tools to analyse it. We started with a brief survey on the history of neuroscience in order to put our contributions in context. Then, we presented the technical background common to our methodological contributions: brain anatomy, DMRI and clustering algorithms (Part I). Next, we proposed two techniques to identify white matter structures in DMRI images. The first of them, groups DMRI image voxels into white matter structures (Part II). The second, clusters fibre tracts obtained through DMRI-based tractography (Part III). For the first algorithm, we used state-of-the-art manifold learning and DMRI techniques. In developing the second, we proposed a novel mathematical framework for white matter fibre bundles. Finally, using this framework we developed two clinical applications: one to perform straightening of the spinal cord using DMRI based tractography (chapter 8); and one to identify group differences in white matter fibre bundles. We applied the latter to detect characteristic differences in schizophrenia (chapter 9). In summary, the contributions of this thesis range from the automatic identification of white matter structures to its statistical analysis to detect characteristic differences of neurological disorders.

Along this thesis, we have tried to make the good mathematical and algorithmic choices to model and solve the problems of interest. This enabled us to propose adequate and efficient algorithms such that we could finally tackle our objective. Briefly, the important and novel methodological contributions of this thesis are:

In Part II, an algorithm to obtain white matter fibre bundles from HARDI images.

This is done by clustering voxels in which the diffusion is represented as ODFs.

In developing this algorithm we:

1. Developed a technique to perform seamless integration of diffusivity and spatial features of the image voxels through Markovian relaxation.
2. Applied machine learning techniques in order to infer the manifold where the voxels of this image lie and improve subsequent statistical analyses.
3. Used Fiedler's contributions to graph theory in order to identify the number of bundles in a given image.
4. Performed clustering of several white matter bundles in an image without the need of user initialization.

In Part III, we developed a mathematical framework for white matter fibre bundles based on Gaussian processes. This contribution has several statistical applications in automatic white matter structure identification and statistical comparison:

1. Seamless parametric representation of individual white matter fibres bundles. This representation retains the smoothness and continuity of these

- anatomical structures.
2. Development of operations among fibre bundles: one to combine them, one to interpolate them and one to quantify their similarity. Moreover, we showed that these operations constitute an inner product space allowing optimizations and statistical analyses.
 3. Calculation of tract probability maps. These maps quantify the probability that a given location in space belongs to a fibre bundle.
 4. The use of the aforementioned characteristics and volumetric atlas in order to automatically identify white matter structures.
 5. Effective identification of white matter structures in a database of 68 registered subject and analysis of the variability of these structures across them.
 6. Application of this mathematical framework to perform spinal cord straightening.
 7. Development of statistical analysis algorithms and their application to detect differences in schizophrenia.

We are confident that these contributions, specially the ones in the third part of this thesis, meet our initial goal; propose a technique to automatically identify human brain's white matter structures in DMRI.

Although our contributions can undoubtedly still be improved, we believe they can benefit to a wide range of clinical and neuroscience studies. Particularly, through the statistical analysis of white matter fibre bundles. In these studies, the proper automatic identification of white matter structures is of utmost importance, specially on those which use large DMRI databases. In these cases, the manual dissection of a white matter structure can be tedious and error prone. This is a consequence of the dependence of the results on the parameters and the human interaction required to perform the dissection. In order to overcome these problems, we introduce in this thesis a tract querying algorithm (Part III). Such algorithm performs automatic identification of white matter structures using anatomical knowledge and results obtained from dense full-brain tractography. Due to the fact that this algorithm is based on a solid mathematical framework, that it requires minimal parameter tuning and it is based on a dense sampling of the fibres, the resulting dissected white matter structures are more robust and reliable than the ones obtained by previous techniques. In what concerns to the detection of biomarkers, it is possible to extend the presented application on schizophrenia to a larger database or to different neurological disorders. In doing so, changes induced by these disorders on specific white matter bundles could be characterized by performing statistical analysis on them. Our mathematical framework could also be used to define a bundle registration algorithm improving the recent work done by [Durrleman et al. \[2009\]](#). Hence, providing the means to quantify their variation in shape and volume. Even more, it is possible

to use this framework in order to perform statistics on shapes where the volume of the overlapping constitutes a significant measure of similarity.

Perspectives on Studies Based on DMRI-tractography

To conclude, we give a perspective on the role of DMRI-based tractography in neuroscience and clinical studies.

A central part of this thesis was about how to integrate the diffusivity and spatial information on each voxel of the images in order to perform their subsequent analysis. In doing so, we found that the fibre representation is more suited to identify the white matter structures than the individual voxels. Tractography integrates diffusivity and spatial information by means of a biologically-oriented physical model. This provides more accurate depiction of the underlying anatomy than individual voxels. However, the accuracy and significance of the fibres obtained through this technique is currently in discussion. On one side, it is argued that tractography is very limited as a tool to study brain structure. This is due to the resolution of the DMRI image, due to dependence of the results on the parameter of the algorithms used to perform this tracing, and due to its incapacity to detect particular anatomical characteristics like the differentiation of afferent and efferent fibres [Johansen-Berg and Behrens: 2006]. On the other side, DMRI-based tractography has been proved to be reproducible [Wakana et al.: 2007, Ciccarelli et al.: 2003], has successfully characterized white matter differences of several neurological disorders [Ciccarelli et al.: 2008] and there has been recent work in rendering more robustly the white matter structures obtained from tractography (Part III). Therefore, we think that tractography provides a useful tool to analyse the human brain's white matter, identify its structure and characterize neurological disorders.

Recent research contributions will also provide a better support for DMRI-tractography based studies. There is a large and exciting amount of research that is currently being performed in order to improve tractography results and reduce its dependence on algorithmic parameters [Descoteaux et al.: 2009a, Qazi et al.: 2009, Fillard et al.: 2009b, Malcolm et al.: 2010]. These new techniques along current developments in the diffusion model within a voxel [Tristán-Vega et al.: 2009, Aganj et al.: 2009, Ghosh and Deriche: 2009, Descoteaux et al.: 2009b], new MRI scanners and acquisition protocols [Sarlls and Pierpaoli: 2009, Aso et al.: 2009b] and new scalar indices which quantify tissue microstructure [Barazany et al.: 2009, Assaf et al.: 2008, Fritzsche et al.: 2010] should not only open but enlarge the doors to new and exciting DMRI-based research of central nervous system connectivity.

Conclusion(français)

Nous avons cherché dans cette thèse à proposer des méthodes permettant d'analyser la structure de la matière blanche du cerveau humain, par le biais de l'Imagerie par Résonance Magnétique de Diffusion (IRMD). Nous nous sommes plus précisément intéressés à développer une technique de dissection *in vivo* et automatique de la matière blanche, ainsi qu'à développer des outils d'analyse statistique pour les structures ainsi mises en évidence. Nous avons en premier lieu présenté un historique rapide des neurosciences, afin de rappeler le contexte dans lequel se situent nos contributions. Nous avons ensuite présenté un diaporama du paysage technique et méthodologique commun à nos contributions : anatomie cérébrale, IRMD et algorithmes de regroupement de faisceaux (Partie I). Nous avons ensuite proposé deux techniques afin d'identifier les structures de la matière blanche dans les images IRMD. La première technique consiste à regrouper des voxels d'une image IRMD en paquets pour retrouver des structures dans la matière blanche (Partie II). La seconde technique quant à elle regroupe des fibres, obtenues par tractographie IRMD, en faisceaux (Partie III). Le premier algorithme utilise des techniques récentes d'apprentissage de variétés et de traitement IRMD. En développant la seconde méthode, nous avons proposé un cadre mathématique nouveau pour l'analyse des faisceaux de fibres de la matière blanche. Ce cadre nous a permis de développer deux applications cliniques : la première consiste à redresser la moëlle épinière à partir de tractographie en IRMD ; la seconde permet d'identifier des différences de faisceaux entre groupes d'individus. Cette deuxième technique a été appliquée pour détecter des différences caractéristiques chez les patients atteints de schizophrénie (Chapitre 9). En résumé, les contributions de cette thèse vont de l'identification des structures de la matière blanche à leur analyse statistique afin de détecter des différences caractéristiques de certaines anomalies neurologique. Nous nous sommes efforcés à faire des choix mathématiques et algorithmiques judicieux pour modéliser et résoudre le problème auquel nous nous intéressions. Cela nous a permis de proposer des algorithmes adaptés et efficaces, si bien que nous sommes parvenus à atteindre nos objectifs. Pour récapituler, les contributions majeures et originales de cette thèse sont :

Dans la Partie II un algorithme pour retrouver les faisceaux de fibres à partir d'images IRMD à haute résolution angulaire. Nous regroupons pour cela des voxels, où la diffusivité est décrite par une fonction de distribution d'orientation. Pour développer cet algorithme, nous avons :

1. Mis en oeuvre une technique pour intégrer de façon transparente la diffusivité et les informations spatiales des voxels de l'image par relaxation de Markov.
2. Appliqué des techniques d'apprentissage pour inférer la variété dans laquelle sont définis les voxels de l'image, et améliorer l'analyse statistique qui en dépend.

3. Utilisé les travaux de Fiedler en théorie des graphes pour identifier le nombre de faisceaux dans une image donnée.
4. Réalisé le regroupement des fibres en plusieurs faisceaux de la matière blanche dans une image, sans besoin d'initialisation par l'utilisateur.

Dans la Partie III, nous avons développé un cadre mathématique pour l'étude des faisceaux de fibres de la matière blanche basé sur la notion de processus gaussien. Cette contribution a plusieurs applications statistiques pour l'identification automatique et la comparaison statistique des structures de la matière blanche :

1. Représentation paramétrique de chaque fibre de la matière blanche. Cette représentation conserve les propriétés de continuité et de régularité de ces structures anatomiques.
2. Définitions d'opérations sur les fibres : combinaison, interpolation et mesure de similarité entre deux fibres. Nous avons par ailleurs montré que ces opérateurs définissent un espace à produit interne, ce qui ouvre la porte à l'optimisation et l'analyse statistique.
3. Calcul de cartes de probabilité d'appartenance à un faisceau. Ces cartes représentent la probabilité qu'un point donné de l'espace appartienne à un faisceau de fibre.
4. Utilisation des caractéristiques citées plus haut et d'un atlas volumétrique pour identifier de façon automatique les structures de la matière blanche.
5. Identification avec succès des structures de la matière blanche dans une base de donnée de 68 sujets recalés, et analyse de la variabilité de ces structures entre sujets.
6. Application de ce cadre mathématique pour redresser la moëlle épinière.
7. Développement d'algorithmes d'analyse statistique et application à la détection de différences en schizophrénie.

Nous sommes convaincus que ces travaux, en particulier ceux présentés dans la troisième partie de cette thèse, remplissent l'objectif que nous nous étions fixés : proposer un algorithme qui permette d'identifier de façon automatique les structures anatomiques de la matière blanche dans le cerveau humain. Bien que, sans nul doute, nos contributions puissent être encore améliorées, nous croyons qu'elles peuvent servir un grand panel d'études cliniques et neurologiques, en particulier grâce à l'analyse statistique des faisceaux de fibres de la matière blanche. Pour ces applications en effet, être capable d'identifier de façon fiable et automatique les structures de la matière blanche est d'une importance capitale, en particulier pour les études sur des grandes bases de données. Dans ce cas, la dissection manuelle d'une structure de la matière blanche peut être délicate, et sujette à erreurs. Les résultats

dépendent en effet des paramètres, et de l'action humaine nécessaire pour faire la dissection. Afin de dépasser ces problèmes, nous introduisons dans cette thèse un algorithme de recherche de faisceau (Partie III). Cet algorithme réalise une identification automatique des structures de la matière blanche à partir de connaissances anatomiques et des résultats de la tractographie exhaustive des fibres de la matière blanche sur le cerveau entier. Étant donné que cet algorithme se base sur un cadre mathématique solide, qu'il nécessite peu d'ajustements de paramètres, et qu'il se base sur un échantillonnage fin des fibres, il permet de disséquer de façon plus robuste et fiable les structures de la matière blanche, en comparaison aux techniques précédentes.

Perspectives sur les Études à base de tractographie en IRMD

En conclusion, nous donnons un point de vue sur le rôle de la tractographie à base d'IRMD en neurosciences et pour les études cliniques. Une partie importante de cette thèse a été consacrée au problème de l'intégration des informations de diffusivité et des informations spatiales en chaque voxel des images pour ainsi pouvoir les analyser. Nous nous sommes aperçus que la représentation sous forme de fibre est mieux appropriée que le simple voxel à l'identification des structures de la matière blanche. La tractographie intègre la diffusivité et la géométrie en utilisant un modèle physique orienté biologie, cela donne une description plus juste de l'anatomie sous-jacente que les voxels considérés individuellement. Cependant, la précision et l'interprétation des fibres obtenues à l'aide de cette technique sont actuellement remises en question. Certains avancent que la tractographie est un outil très limité pour l'étude de la structure du cerveau ; cela est attribué à la résolution des images IRMD, au choix déterminant des paramètres pour le résultat des algorithmes de tractographie, ainsi qu'à l'impossibilité de discerner certaines particularités anatomiques, comme par exemple discriminer une fibre afférente d'une fibre efférente [[Johansen-Berg and Behrens: 2006](#)]. D'un autre côté, il a été montré que la tractographie IRMD est reproductible [[Wakana et al.: 2007](#), [Ciccarelli et al.: 2003](#)], a permis d'identifier avec succès plusieurs anomalies de la matière blanche caractéristiques de certaines pathologies neurologiques [[Ciccarelli et al.: 2008](#)], et des efforts récents ont été mis en oeuvre pour rendre l'extraction de structures de la matière blanche à partir de tractographie plus robuste (Partie III). Ainsi nous sommes convaincus que la tractographie apporte un excellent outil pour l'analyse de la matière blanche du cerveau humain, l'identification de ses structures et la caractérisation des anomalies neurologiques. Les travaux de recherche les plus récents vont également apporter une aide supplémentaire pour les études basées sur la tractographie en IRMD. Il y a un nombre toujours plus important et encourageant de recherches en cours afin d'améliorer les résultats de la tractographie, et de les rendre au maximum indépendants des paramètres des algorithmes [[Descoteaux et al.: 2009a](#), [Qazi et al.: 2009](#), [Fillard et al.: 2009b](#), [Malcolm et al.: 2010](#)]. Ces récentes techniques,

les développements actuels de modèles décrivant la diffusion dans un voxel [Tristán-Vega et al.: 2009, Aganj et al.: 2009, Ghosh and Deriche: 2009, Descoteaux et al.: 2009b], l'arrivée de nouveaux scanners IRM et protocoles d'acquisition [Sarlls and Pierpaoli: 2009, Aso et al.: 2009b], ainsi que le développement de nouveaux indices scalaires qui décrivent de façon quantitative la microstructure des tissus [Barazany et al.: 2009, Assaf et al.: 2008, Fritzsche et al.: 2010] devraient non seulement ouvrir, mais également agrandir les portes vers de nouvelles recherches de la connectivité dans le système nerveux central basées sur la tractographie en IRMD.

Appendix

PUBLICATIONS OF THE AUTHOR

JOURNAL PUBLICATIONS

- D. Wassermann**, L. Bloy, E. Kanterakis, R. Verma, and R. Deriche. Unsupervised white matter fiber clustering and tract probability map generation: Applications of a gaussian process framework for white matter fibers. *NeuroImage*, 51, 228-241, 2010. DOI: 10.1016/j.neuroimage.2010.01.004.
- C. Lenglet, J. Campbell, M. Descoteaux, G. Haro, P. Savadjiev, **D. Wassermann**, A. Anwender, R. Deriche, B. Pike, G. Sapiro, K. Siddiqi, and P. Thompson. Mathematical methods for diffusion MRI processing. *Neuroimage*, 45(1):S111–S122, 2009. DOI:10.1016/j.neuroimage.2008.10.054.
- C. Delmaire, M. Vidailhet, **D. Wassermann**, M. Descoteaux, R. Valabregue, F. Bourdain, C. Lenglet, S. Sangla, A. Terrier, R. Deriche, and S. Lehericy. Diffusion abnormalities in the primary sensorimotor pathways in writer’s cramp. *Arch Neurol.*, 66(4):502–508, 2009.
- D. Wassermann**, M. Descoteaux, and R. Deriche. Diffusion maps clustering for magnetic resonance Q-Ball imaging segmentation. *International J Biomedical Imaging*, 2008, 2008. DOI: 10.1155/2008/526906

CONFERENCES AND WORKSHOPS WITH PROCEEDINGS

- D. Wassermann**, E. Kanterakis, R. C. Gur, R. Deriche, and R. Verma. Diffusion-based population statistics using tract probability maps. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2010.
- D. Wassermann**, J. Cohen-Adad, S. Lehericy, H. Benali, S. Rossignol, and R. Deriche. Straightening the spinal cord using fiber tractography. In *International Symposium on Biomedical Imaging*, 2010.

- D. Wassermann**, L. Bloy, R. Verma, and R. Deriche. Bayesian framework for white matter fiber similarity measure. In International Symposium on Biomedical Imaging, Boston, USA, 2009
- D. Wassermann** and R. Deriche. Simultaneous manifold learning and clustering: Grouping white matter fiber tracts using a volumetric white matter atlas. In MICCAI Workshops, 2008.
- D. Wassermann**, L. Bloy, R. Verma, and R. Deriche. A Gaussian Process based framework for white matter fiber tracts and bundles, applications to fiber clustering. In MICCAI Workshops, 2009.
- D. Wassermann**, M. Descoteaux, and R. Deriche. Diffusion maps segmentation of magnetic resonance Q-Ball imaging. In IEEE Workshop on Mathematical Methods in Biological Imaging, 2007.

CONFERENCE ABSTRACTS

- D. Wassermann**, J. Cohen-Adad, S. Lehericy, H. Benali, S. Rossignol, and R. Deriche. Fully automated straightening of the spinal cord using fiber tractography. In 18th. International Symposium on Magnetic Resonance in Medicine, 2010.
- C. Delmaire, M. Vidailhet, M. Descoteaux, **D. Wassermann**, F. Bourdain, C. Lenglet, S. Sangla, A. Terrier, R. Deriche, and S. Lehericy. Diffusion tensor imaging of white matter abnormalities in patients with writer's cramp. In 16th. International Symposium on Magnetic Resonance in Medicine, 2008.
- D. Wassermann** and R. Deriche. Avoiding artifacts in spectral white matter fiber clustering and embedding. In Proceedings of the 13th Organization for Human Brain Mapping Annual Meeting, volume 36 of Neuroimage. Organization for the Human Brain Mapping, 2007.

USEFUL MATHEMATICAL PROPERTIES

B.1 PROPERTIES OF GAUSSIAN DISTRIBUTIONS _____

B.1.1 Condition and Marginalization

Let \mathbf{x} and \mathbf{y} be jointly Gaussian random variables

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathcal{G} \left(\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} A & C \\ C^T & B \end{pmatrix} \right) = \mathcal{G} \left(\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \tilde{A} & \tilde{C} \\ \tilde{C}^T & \tilde{B} \end{pmatrix}^{-1} \right) \quad (\text{B.1})$$

Then, the following properties hold

$$\mathbf{x} \sim \mathcal{G}(\boldsymbol{\mu}_x, A) \quad (\text{B.2})$$

$$\mathbf{x}|\mathbf{y} \sim \mathcal{G}(\boldsymbol{\mu}_x + CB^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), A - CB^{-1}C^T) \quad (\text{B.3})$$

$$\mathbf{x}|\mathbf{y} \sim \mathcal{G}(\boldsymbol{\mu}_x + \tilde{A}^{-1}\tilde{C}(\mathbf{y} - \boldsymbol{\mu}_y), \tilde{A}^{-1}) \quad (\text{B.4})$$

B.2 LINEAR ALGEBRA _____

Partitioned inverse equation

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{pmatrix} \quad (\text{B.5})$$

Bibliography

- Aganj, I., Lenglet, C., Sapiro, G., 2009. Accurate ODF Reconstruction in Q-ball Imaging. *Neuroimage* 47, 51–51.
- Alexander, D. C., Barker, G. J., Arridge, S. R., Aug 2002. Detection and modeling of non-gaussian apparent diffusion coefficient profiles in human brain data. *Magn Reson Med* 48 (2), 331–340.
- Alexander, D. C., Jiang, T., Westin, C.-F., nov 2007. Guest editorial special issue on computational diffusion mri. *Medical Imaging, IEEE Transactions on* 26 (11), 1425–1427.
- Anderson, A., 2005. Measurements of fiber orientation distributions using high angular resolution diffusion imaging. *Magnetic Resonance in Medicine* 54, 1194–1206.
- Anwander, A., Tittgemeyer, M., von Cramon, D., Friederici, A., Knösche, T., apr 2007. Connectivity-based parcellation of broca’s area. *Cerebral Cortex* 17, 16–825.
- Aristotle, 1942. *Generation of Animals*. Harvard.
- Aristotle, 1955. *Parts of Animals*. Harvard.
- Aristotle, 1957a. *On Sense and Sensible Objects in Parva Naturalia*. Harvard.
- Aristotle, 1957b. *On Sleep and Waking in Parva Naturalia*. Harvard.
- Aristotle, 1957c. *On youth and old ange in Parva Naturalia*. Harvard.
- Arsigny, V., Fillard, P., Pennec, X., Ayache, N., August 2006. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *MRM* 56 (2), 411–421.
- Ashtari, M., Cottone, J., Ardekani, B., Cervellione, K., Szeszko, P., Wu, J., Chen, S., Kumra, S., 2007. Disruption of white matter integrity in the inferior longitudinal fasciculus in adolescents with schizophrenia as revealed by fiber tractography. *Arch of Gen Psych* 64 (11), 1270.
- Aso, T., Urayama, S., Poupon, C., Sawamoto, N., Fukuyama, H., Bihan, D., 2009a. An intrinsic diffusion response function for analyzing diffusion functional MRI time series. *Neuroimage* 47 (4), 1487–1495.

- Aso, T., Wiggins, C., Poupon, C., Urayama, S., Fukuyama, H., Le Bihan, D., 2009b. Diffusion-weighted fMRI at 7T: a comparison with 3T. *Neuroimage* 47, 186–186.
- Assaf, Y., Blumenfeld-Katzir, T., Yovel, Y., Basser, P., 2008. AxCaliber: a method for measuring axon diameter distribution from diffusion MRI. *Magnetic Resonance in Medicine* 59 (6), 1347–1354.
- Audoin, B., Guye, M., Reuter, F., Au Duong, M., Confort-Gouny, S., Malikova, I., Soulier, E., Viout, P., Chérif, A., Cozzone, P., et al., 2007. Structure of WM bundles constituting the working memory system in early multiple sclerosis: a quantitative DTI tractography study. *Neuroimage* 36 (4), 1324–1330.
- Awate, S., Hui, Z., Gee, J., nov 2007. A fuzzy, nonparametric segmentation framework for dti and mri analysis: With applications to dti-tract extraction. *Medical Imaging, IEEE Transactions on* 26 (11), 1525–1536.
- Ayache, N., Ourselin, S., Maeder, A. (Eds.), 2007. MICCAI. No. 4791 in LNCS. Springer Berlin / Heidelberg.
- Ball, G., Hall, D., 1965. ISODATA, a novel method of data analysis and pattern classification. Tech. rep., Stanford University.
- Barazany, D., Basser, P., Assaf, Y., 2009. In vivo measurement of axon diameter distribution in the corpus callosum of rat brain. *Brain* 132 (5), 1210–1221.
- Basser, P., 1995. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR in Biomedicine* 8 (7), 333–344.
- Basser, P., Mattiello, J., Le Bihan, D., 1994a. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance-Series B* 103 (3), 247–254.
- Basser, P., Mattiello, J., Le Bihan, D., 1994b. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal* 66 (1), 259–267.
- Basser, P., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A., 2000. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine* 44 (4), 625–632.
- Basser, P., Pierpaoli, C., 1996. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Mag Res B* 111 (3), 209–219.
- Batchelor, P. G., Calamante, F., Tournier, J.-D., Atkinson, D., Hill, D. L. G., Connelly, A., 2006. Quantification of the shape of fiber tracts. *MRM* 55 (4), 894–903.
- Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system- a technical review. *NMR in Biomedicine* 15 (7-8), 435–455.

- Beaulieu, C., 2009. Diffusion MRI, Ch. The Biological Basis of Diffusion Anisotropy. In: [Johansen-Berg and Behrens: 2009].
- Behrens, T. E. J., Johansen-Berg, H., Woolrich, M. W., Smith, S. M., Wheeler-Kingshott, C. A. M., Boulby, P. A., Barker, G. J., Sillery, E. L., Sheehan, K., Ciccarelli, O., 2003. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience* 6, 750–757.
- Belkin, M., aug 2003. Problems of learning on manifolds. Ph.D. thesis, University of Chicago.
- Belkin, M., Niyogi, P., jun 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15 (6), 1373–1396.
- Bernstein, M., Silva, V., Langford, J. C., Tenenbaum, J. B., dec 2000. Graph approximations to geodesics on embedded manifolds.
- Blei, D., Griffiths, T., Jordan, M., Tenenbaum, J., 2004. Hierarchical topic models and the nested Chinese restaurant process. In: *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference*. Bradford Book, p. 17.
- Blei, D., Ng, A., Jordan, M., 2003. Latent dirichlet allocation. *The Journal of Machine Learning Research* 3, 993–1022.
- Bloch, F., Hansen, W., Packard, M., 1946. Nuclear induction. *Physical Review* 69 (3-4), 127.
- Breasted, J., 1930. *The Edwin Smith Surgical Papyrus*. University of Chicago Press.
- Broca, P., 1861. Remarques sur le siege de la faculté du langage articulé, suivies d’une observation d’aphemie (Perte de la Parole). *Bulletin de la Société Anatomique* 6, 330–357.
- Brown, R., 1828. XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine Series 2* 4 (21), 161–173.
- Buhmann, M., 1998. Radial functions on compact support. *Proceedings of the Edinburgh Mathematical Society* 41, 33–46.
- Buhmann, M., 2004. *Radial Basis Functions*. Cambridge Univ Press.
- Bullock, T., Bennett, M., Johnston, D., Josephson, R., Marder, E., Fields, R., 2005. The neuron doctrine, redux. *Science* 310 (5749), 791–793.

- Bürgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J. M., Zilles, K., 2006. White matter fiber tracts of the human brain: Three-dimensional mapping at microscopic resolution, topography and intersubject variability. *NeuroImage* 29 (4), 1092 – 1105.
- Burns, J., Job, D., Bastin, M. E., Whalley, H., Macgillivray, T., Johnstone, E. C., Lawrie, S. M., 2003. Structural disconnectivity in schizophrenia: a diffusion tensor magnetic resonance imaging study. *The British journal of psychiatry: the journal of mental science* 182, 439.
- Callaghan, P., Eccles, C., Xia, Y., 1988. Nmr microscopy of dynamic displacements: k-space and q-space imaging. *Journal of Physics E: Scientific Instruments* 21, 820–822.
- Campbell, J., Siddiqi, K., Rymar, V., Sadikot, A., Pike, G., 2005. Flow-based fiber tracking with diffusion tensor and q-ball data: validation and comparison to principal diffusion direction techniques. *NeuroImage* 27 (4), 725–736.
- Carr, H., Purcell, E., 1954. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review* 94 (3), 630–638.
- Carreira-Perpiñán, M., 2000. Mode-Finding for Mixtures of Gaussian Distributions. *Pattern Analysis and Machine Intelligence*, 1318–1323.
- Carreira-Perpinan, M., May 2007. Gaussian mean-shift is an em algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29 (5), 767–776.
- Castaño-Moraga, C., Lenglet, C., Deriche, R., Ruiz-Alzola, J., 2007. A Riemannian approach to anisotropic filtering of tensor fields. *Signal Processing* 87 (2), 263–276.
- Catani, M., 2006. Diffusion tensor magnetic resonance imaging tractography in cognitive disorders. *Current opinion in neurology* 19 (6), 599.
- Catani, M., Allin, M., Husain, M., Pugliese, L., Mesulam, M., Murray, R., Jones, D., 2007. Symmetries in human brain language pathways correlate with verbal recall. *Proceedings of the National Academy of Sciences* 104 (43), 17163.
- Catani, M., de Schotten, M. T., 2008. A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex* 44 (8), 1105 – 1132, special Issue on "Brain Hodology - Revisiting disconnection approaches to disorders of cognitive function".
- Catani, M., Howard, R., Pajevic, S., Jones, D., 2002. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage* 17 (1), 77–94.
- Catani, M., Jones, D., Donato, R., 2003. Occipito-temporal connections in the human brain. *Brain* 126 (9), 2093.

- Celeux, G., Soromenho, G., 1996. An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification* 13 (2), 195–212.
- Cercignani, M., Horsfield, M. A., 2001. The physical basis of diffusion-weighted mri. *Journal of the Neurological Sciences* 186 (Supplement 1), S11 – S14.
- Cheng, Y., 1995. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17 (8), 790–799.
- Cho, S., Kim, D., Kim, D., Kim, Y., Lee, C., Jang, S., 2007. Motor outcome according to the integrity of the corticospinal tract determined by diffusion tensor tractography in the early stage of corona radiata infarct. *Neuroscience letters* 426 (2), 123–127.
- Chung, F., 1997. Spectral graph theory. CBMS-AMS.
- Ciccarelli, O., Catani, M., Johansen-Berg, H., Clark, C., Thompson, A., 2008. Diffusion-based tractography in neurological disorders: concepts, applications, and future developments. *Lancet Neurology* 7 (8), 715–727.
- Ciccarelli, O., Parker, G., Toosy, A., Wheeler-Kingshott, C., Barker, G., Boulby, P., Miller, D., Thompson, A., 2003. From diffusion tractography to quantitative white matter tract measures: a reproducibility study. *Neuroimage* 18 (2), 348–359.
- Ciccarelli, O., Toosy, A., Hickman, S., Parker, G., Wheeler-Kingshott, C., Miller, D., Thompson, A., 2005. Optic radiation changes after optic neuritis detected by tractography-based group mapping. *Human brain mapping* 25 (3), 308–316.
- Clarke, E., 1970. The history of the neurological sciences. *Proceedings of the Royal Society of Medicine* 63 (1), 21.
- Cohen-Adad, J., Benali, H., Hoge, R., Rossignol, S., 2008. In vivo DTI of the healthy and injured cat spinal cord at high spatial and angular resolution. *Neuroimage* 40 (2), 685–697.
- Coifman, R., Lafon, S., Lee, A., Maggioni, M., Nadler, B., Warner, F., Zucker, S., may 2005. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences* 102 (21), 7426–7431.
- Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. *IEEE PAMI* 24 (5), 603–619.
- Concha, L., Gross, D., Wheatley, B., Beaulieu, C., 2006. Diffusion tensor imaging of time-dependent axonal and myelin degradation after corpus callosotomy in epilepsy patients. *Neuroimage* 32 (3), 1090–1099.

- Conturo, T., Lori, N., Cull, T., Akbudak, E., Snyder, A., Shimony, J., McKinstry, R., Burton, H., Raichle, M., 1999. Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences* 96 (18), 10422–10427.
- Corouge, I., Fletcher, P. T., Joshi, S., Gouttard, S., Gerig, G., 2006. Fiber tract-oriented statistics for quantitative diffusion tensor mri analysis. *MIA* 10 (5), 786–798.
- Cowan, W., Gottlieb, D., Hendrickson, A., Price, J., Woolsey, T., 1972. The autoradiographic demonstration of axonal connections in the central nervous system. *Brain Research* 37 (1), 21.
- Cremers, D., Rousson, M., Deriche, R., 2007. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *IJCV* 72 (2), 195–215.
- Cressie, N., 1991. *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics Section (EUA).
- Crivellato, E., Ribatti, D., 2007. Soul, mind, brain: Greek philosophy and the birth of neuroscience. *Brain Research Bulletin* 71 (4), 327 – 336.
- Damon, B., Ding, Z., Anderson, A., Freyer, A., Gore, J., 2002. Validation of diffusion tensor MRI-based muscle fiber tracking. *Magnetic Resonance in Medicine* 48 (1), 97–104.
- Dauguet, J., Peled, S., Berezovskii, V., Delzescaux, T., Warfield, S. K., Born, R., Westin, C.-F., Aug 2007. Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain. *Neuroimage* 37 (2), 530–538.
- De Groot, M., 2004. *Optimal Statistical Decisions*. Classics Library. Wiley.
- Dejerine, J., Dejerine-Klumpke, A., 1895. *Anatomie des centres nerveux*. Rueff Paris.
- Dempster, A., Laird, N., Rubin, D., et al., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1), 1–38.
- Deriche, R., Descoteaux, M., April 2007. Splitting tracking through crossing fibers: multidirectional fiber assignment by continuous tracking (M-FACT) assignment by continuous tracking (M-FACT). In: *International Symposium on Biomedical Imaging*. IEEE.
- Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R., 2007a. Regularized, fast and robust analytical q-ball imaging. *MRM* 58 (3), 497–510.
- Descoteaux, M., Deriche, R., 2007. Segmentation of q-ball images using statistical surface evolution. In: [Ayache et al.: 2007].

- Descoteaux, M., Deriche, R., 2008. High angular resolution diffusion mri segmentation using region-based statistical surface evolution. *J Mathetical Imaging in Vision*(in press).
- Descoteaux, M., Deriche, R., Knösche, T., Anwander, A., 2009a. Deterministic and probabilistic tractography based on complex fiber orientation distributions. *Trans. in Med. Imag.* 28 (2), 269–286.
- Descoteaux, M., Deriche, R., Le Bihan, D., Mangin, J., Poupon, C., 2009b. Diffusion Propagator Imaging: Using Laplace’s Equation and Multiple Shell Acquisitions to Reconstruct the Diffusion Propagator. In: *Proc. of 21st Intl. Conf. on IPMI*. Springer.
- Descoteaux, M., Savadjiev, P., Campbell, J., Pike, G., Siddiqi, K., Deriche, R., April 2007b. Validation and comparison of analytical Q-ball imaging methods. In: *International Symposium on Biomedical Imaging*. IEEE, pp. 1084–1087.
- Dhillon, I. S., Guan, Y., Kulis, B., nov 2007. Weighted graph cuts without eigenvectors a multilevel approach. *Transactions on Pattern Analysis and Machine Intelligence* 29 (11), 1944–1957.
- Ding, Z., Gore, J., Anderson, A., 2003. Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. *MRM* 49, 716–721.
- Donoho, D., Grimes, C., 2003. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences* 100 (10), 5591–5596.
- Duchon, J., 1977. Splines Minimizing Rotation-Invariant Semi-Norms in Sobolev Spaces. *Constructive Theory of Functions of Several Variables*, 85–100.
- Duda, R., Hart, P., Stork, D., 2001. *Pattern classification*, 2nd Edition. Wiley.
- Dunn, J., 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Cybernetics and Systems* 3 (3), 32–57.
- Durrleman, S., Pennec, X., Trouvé, A., Ayache, N., 2009. Statistical models of sets of curves and surfaces based on currents. *Medical Image Analysis In Press*, Accepted Manuscript, –.
- Einstein, A., 1905. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen:w. *Annalen der physik* 17 (4), 548–560.
- Ellison-Wright, I., Bullmore, E., 2009. Meta-analysis of diffusion tensor imaging studies in schizophrenia. *Schizophrenia Research* 108 (1-3), 3–10.

- Ennis, D. B., Kindlmann, G., Jan 2006. Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. *Magn Reson Med* 55 (1), 136–146.
- Ester, M., Kriegel, H., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proc. KDD*. Vol. 96. pp. 226–231.
- Etynger, P., 2007. *Statistical learning, Shape Manifolds and Applications to Image Segmentation*. Ph.D. thesis, École Nationale des Ponts et Chaussées.
- Faulkner, R. O., 1972. *The Ancient Egyptian Book of the Dead*. University of Texas Press.
- Feddern, C., Weickert, J., Burgeth, B., 2003. Level-set methods for tensor-valued images. In: *2nd Workshop on Geometric and Level Set Methods in Computer Vision*. IEEE.
- Fick, A., 1855. Ueber diffusion. *Annalen der Physik* 94, 59–86.
- Fick, A., Müller, J., 1866. *Die medicinische Physik*. F. Vieweg.
- Fiedler, M., 1973. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal* 23 (98), 298–306.
- Fiedler, M., 1975. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal* 25, 619–633.
- Figueiredo, M., Jain, A., 2002. Unsupervised learning of finite mixture models. *IEEE Transactions on pattern analysis and machine intelligence*, 381–396.
- Filippone, M., Camastra, F., Masulli, F., Rovetta, S., 2008. A survey of kernel and spectral methods for clustering. *Pattern Recognition* 41 (1), 176–190.
- Fillard, P., Descoteaux, M., Poupon, C., Mangin, J.-F. (Eds.), 2009a. *MICCAI'09 Fiber Cup*. Laboratoire de Neuroimagerie Assistée par Ordinateur.
- Fillard, P., Pennec, X., Arsigny, V., Ayache, N., 2007. Clinical DT-MRI estimation, smoothing, and fiber tracking with log-Euclidean metrics. *IEEE Transactions on Medical Imaging* 26 (11), 1472–1485.
- Fillard, P., Poupon, C., Mangin, J.-F., 2009b. A novel global tractography algorithm based on an adaptive spin glass model. In: *MICCAI*.
- Fitzsimmons, J., Kubicki, M., Smith, K., Bushell, G., Estepar, R., Westin, C., Nestor, P., Niznikiewicz, M., Kikinis, R., McCarley, R., et al., 2009. Diffusion tractography of the fornix in schizophrenia. *Schizophrenia Research* 107 (1), 39–46.
- Forgy, E., 1965. Cluster analysis of multivariate data: Efficiency vs. interpretability of classifications. *Biometrics* 21 (3), 768.

- Fowlkes, C., Belongie, S., Chung, F., Malik, J., Feb 2004. Spectral grouping using the nystrom method. *IEEE PAMI* 26 (2), 214–225.
- Fraley, C., Raftery, A., 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal* 41 (8), 578.
- Fritzsche, K. H., Laun, F. B., Meinzer, H.-P., Stieltjes, B., 2010. Opportunities and pitfalls in the quantification of fiber integrity: What can we gain from Q-ball imaging? *NeuroImage In Press, Accepted Manuscript*, –.
- FSL, 2006. Fmrib software library. University of Oxford.
- Fuentes, M., Smith, R., 2001. A new class of nonstationary spatial models. *Journal of American Statistical Association*.
- Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory, IEEE Transactions on* 21 (1), 32–40.
- Gall, F., Spurzheim, 1810. *On the Functions of the Brain and of Each of Its Parts*. Marsh, Capen & Lyon.
- Ghosh, A., Deriche, R., may 2009. From second to higher order tensors in diffusion-mri. In: Aja-Fernández, S., de Luis García, R., Tao, D., Li, X. (Eds.), *Tensors in Image Processing and Computer Vision. Advances in Pattern Recognition*. Springer London, Ch. 9.
- Glickstein, M., 2006. Golgi and Cajal: The neuron doctrine and the 100th anniversary of the 1906 Nobel Prize. *Current Biology* 16 (5), 147–151.
- Golgi, C., Bentivoglio, M., Swanson, L., 2001. On the fine structure of the pes Hippocampi major (with plates XIII-XXIII). 1886. *Brain Res Bull* 54 (5), 461–83.
- Goodlett, C. B., Fletcher, P. T., Gilmore, J. H., Gerig, G., 2009. Group analysis of dti fiber tract statistics with application to neurodevelopment. *NeuroImage* 45 (1, Supplement 1), S133 – S142, *mathematics in Brain Imaging*.
- Gössl, C., Fahrmeir, L., Pütz, B., Auer, L., Auer, D., 2002. Fiber tracking from dti using linear state space models: detectability of the pyramidal tract. *Neuroimage* 16 (2), 378–388.
- Gray, H., 1918. *Anatomy of the human body*. Philadelphia: Lea & Febiger, 1918; Bartleby.com, 2000.
- Gross, C., 1995. Aristotle on the brain. *The Neuroscientist* 1 (4), 245.
- Gross, C., 1998a. Galen and the squealing pig. *The Neuroscientist* 4 (3), 216.

- Gross, C., 2007. The discovery of motor cortex and its background. *Journal of the History of the Neurosciences* 16 (3), 320–331.
- Gross, C. G., 1998b. *Brain, Vision, Memory. Tales in the History of Neuroscience*. MIT Press.
- Hageman, N., Shattuck, D., Narr, K., Toga, A., April 2006. A diffusion tensor imaging tractography algorithm based on navier-stokes fluid mechanics. In: *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*. pp. 798–801.
- Hagen, L., Kahng, A., sep 1992. New spectral methods for ratio cut partitioning and clustering. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* 11 (9), 1074–1085.
- Hagmann, P., Jonasson, L., Deffieux, T., Meuli, R., Thiran, J.-P., Wedeen, V. J., aug 2006. Fibertract segmentation in position orientation space from high angular resolution diffusion MRI. *Neuroimage* 32 (2), 665–675.
- Hahn, E. L., Nov 1950. Spin echoes. *Phys. Rev.* 80 (4), 580–594.
- Hansen, M., Yu, B., 2001. Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 746–774.
- Haro, G., Randall, G., Sapiro, G., 2006. Stratification learning: Detecting mixed density and dimensionality in high dimensional point clouds. *Proc. 19th Conf. Adv. Neural Inform. Process. Syst.*
- Hasan, K. M., Iftikhar, A., Kamali, A., Kramer, L. A., Ashtari, M., Cirino, P. T., Papanicolaou, A. C., Fletcher, J. M., Ewing-Cobbs, L., 2009. Development and aging of the healthy human brain uncinata fasciculus across the lifespan using diffusion tensor tractography. *Brain Research* 1276, 67 – 76.
- Hastie, Tibshirani, Friedman, 2009. *The Elements of Statistical Learning*, 3rd Edition. Springer-Verlag.
- Heemskerk, A., Strijkers, G., Vilanova, A., Drost, M., Nicolay, K., 2005. Determination of mouse skeletal muscle architecture using three-dimensional diffusion tensor imaging. *Magnetic Resonance in Medicine* 53 (6), 1333–1340.
- Hein, M., Audibert, J., 2005. Intrinsic dimensionality estimation of submanifolds in R^d . *ACM International Conference Proceeding Series* 119, 289–296.
- Hein, M., Audibert, J., von Luxburg, U., 2007. Graph laplacians and their convergence on random neighborhood graphs. *The J Machine Learning Research* 8, 1325–1370.

- Hess, C., Mukherjee, P., Han, E., Xu, D., Vigneron, D., 2006. Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. *Magnetic Resonance in Medicine* 56, 104–117.
- Higdon, D., Swall, J., Kern, J., 1999. Non-stationary spatial modeling. *Bayesian statistics* 6, 761–768.
- Hinneburg, A., Gabriel, H., 2007. Denclue 2.0: Fast clustering based on kernel density estimation. *Lecture Notes in Computer Science* 4723, 70.
- Hinneburg, A., Keim, D., 1998. An efficient approach to clustering in large multimedia databases with noise. *Knowledge Discovery and Data Mining* 5865.
- Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., Calabresi, P. A., Pekar, J. J., van Zijl, P. C. M., Mori, S., 2008. Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. *NeuroImage* 39 (1), 336–347.
- Iturria-Medina, Y., Canales-Rodriguez, E. J., Melie-Garcia, L., Valdes-Hernandez, P. A., Martinez-Montes, E., Aleman-Gomez, Y., Sanchez-Bornot, J. M., 2007. Characterizing brain anatomical connections using diffusion weighted mri and graph theory. *NeuroImage* 36 (3), 645–660.
- Jaermann, T., Crelier, G., Pruessmann, K., Golay, X., Netsch, T., van Muiswinkel, A., Mori, S., van Zijl, P., Valavanis, A., Kollias, S., et al., 2004. SENSE-DTI at 3T. *Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine/Society of Magnetic Resonance in Medicine* 51 (2), 230.
- Jain, A., 2009. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*.
- Jain, A. K., Dubes, R. C., 1988. *Algorithms for clustering data*. Prentice Hall.
- Jain, A. K., Duin, R. P., Mao, J., 2000. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (1), 4–37.
- Jain, A. K., Law, M. H., dec 2005. Data clustering: A user's dilemma. In: *Pattern Recognition and Machine Intelligence*. No. 3776 in *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, pp. 1–10.
- Jain, A. K., Murty, M. N., Flynn, P., 1999. Data clustering: a review. *ACM Computing Surveys (CSUR)* 31 (3), 264–323.
- Jansons, K., Alexander, D., 2003. Persistent angular structure: new insights from diffusion magnetic resonance imaging data. *Inverse Problems* 19 (5), 1031–1046.
- Jarvis, R., Patrick, E., 1973. Clustering using a similarity measure based on shared near neighbors. *IEEE Transactions on Computers* 22 (11), 1025–1034.

- Jbabdi, S., Bellec, P., Marrelec, G., Perlberg, V., Benali, H., 2004. A level set method for building anatomical connectivity paths between brain areas using DTI. In: International Symposium on Biomedical Imaging. Vol. 1. IEEE, pp. 1024–1027.
- Jbabdi, S., Woolrich, M. W., Andersson, J. L. R., Behrens, T. E. J., Aug 2007. A bayesian framework for global tractography. *Neuroimage* 37 (1), 116–129.
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., October 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17 (2), 825–841.
- Johansen-Berg, H., Behrens, T., 2006. Just pretty pictures? What diffusion tractography can add in clinical neuroscience. *Current opinion in neurology* 19 (4), 379.
- Johansen-Berg, H., Behrens, T. E. J. (Eds.), 2009. *Diffusion MRI*. Academic Press.
- Jonasson, L., Bresson, X., Hagmann, P., Cuisenaire, O., Meuli, R., Thiran, J., 2005. White matter fiber tract segmentation in DT-MRI using geometric flows. *Medical Image Analysis* 9 (3), 223–236.
- Jonasson, L., Bresson, X., Hagmann, P., Thiran, J.-P., Wedeen, V., Feb 2006. Representing diffusion mri in 5d simplifies regularization and segmentation of white matter tracts, submitted to *For IEEE Transactions in Medical Imaging*.
- Jonasson, L., Hagmann, P., Pollo, C., Bresson, X., Wilson, C., Meuli, R., Thiran, J., 2007. A level set method for segmentation of the thalamus and its nuclei in DT-MRI. *Signal Processing* 87, 309–321.
- Jones, D., Horsfield, M., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine* 42 (3), 515–525.
- Jones, D. K., 2008. Studying connections in the living human brain with diffusion mri. *Cortex* 44 (8), 936 – 952, special Issue on Brain Hodology - Revisiting disconnection approaches to disorders of cognitive function”.
- Jones, D. K., Williams, S. C. R., Gasston, D., Horsfield, M. A., Simmons, A., Howard, R., 2002. Isotropic resolution diffusion tensor imaging with whole brain acquisition in a clinically acceptable time. *Human Brain Mapping* 15, 216–230.
- Jørgensen, J., 1988. Trepanation as a therapeutic measure in ancient (pre-Inka) Peru. *Acta neurochirurgica* 93 (1), 3–5.
- Karypis, G., Han, E., Kumar, V., 1999. Chameleon: A hierarchical clustering algorithm using dynamic modeling. *IEEE computer* 32 (8), 68–75.
- Kaufman, L., Rousseeuw, P., 1987. Clustering by means of medoids. Tech. rep., Technische Hogeschool, Delft(Netherlands). Dept. of Mathematics and Informatics.

- Kindlmann, G., Estepar, R., Smith, S., Westin, C., 2009. Sampling and Visualizing Creases with Scale-Space Particles. *IEEE Trans. on Viz. and Comp. Graph.* 15 (6), 1415–1424.
- Kindlmann, G., Estépar, R. S. J., Niethammer, M., Haker, S., Westin, C.-F., 2007. Geodesic-loxodromes for diffusion tensor interpolation and difference measurement. In: [Ayache et al.: 2007].
- Kolmogorov, A. N., 1956. *Foundations of the Theory of Probability*. Chelsea.
- Konishi, J., Yamada, K., Kizu, O., Ito, H., Sugimura, K., Yoshikawa, K., Nakagawa, M., Nishimura, T., 2005. MR tractography for the evaluation of functional recovery from lenticulostriate infarcts. *Neurology* 64 (1), 108.
- Kubicki, M., McCarley, R., Westin, C., Park, H., Maier, S., Kikinis, R., Jolesz, F., Shenton, M., 2007. A review of diffusion tensor imaging studies in schizophrenia. *J Psych Res* 41 (1-2), 15–30.
- Kubicki, M., Shenton, M. E., 2009. Diffusion MRI, Ch. Diffusion Tensor Imaging and Its Application to Schizophrenia and Related Disorders. In: [Johansen-Berg and Behrens: 2009].
- Lafon, S., 2004. Diffusion maps and geometric harmonics. Ph.D. thesis, Yale University.
- Lafon, S., Lee, A., Sep 2006. Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE PAMI* 28 (9), 1393–1403.
- Lauterbur, P., 1973. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature* 242 (5394), 190–191.
- Lawes, I., Barrick, T., Murugam, V., Spierings, N., Evans, D., Song, M., Clark, C., 2008. Atlas-based segmentation of white matter tracts of the human brain using diffusion tensor tractography and comparison with classical dissection. *NeuroImage* 39 (1), 62–79.
- Lazar, M., Alexander, A. L., 2003. An error analysis of white matter tractography methods: synthetic diffusion tensor field simulations. *NeuroImage* 20 (2), 1140 – 1153.
- Le Bihan, D., Breton, E., 1985. Imagerie de diffusion in vivo par résonance magnétique nucléaire. *Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre* 301 (15), 1109–1112.

- Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet, M., 1986. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology* 161 (2), 401–407.
- Le Bihan, D., Urayama, S.-i., Aso, T., Hanakawa, T., Fukuyama, H., 2006. Direct and fast detection of neuronal activation in the human brain with diffusion mri. *Proceedings of the National Academy of Sciences* 103 (21), 8263–8268.
- Lee, A., Wasserman, L., 2009. Spectral connectivity analysis. Submitted for publication (preprint: arXiv/0811.0121).
- Leemans, A., Sijbers, J., Backer, S. D., Vandervliet, E., Parizel, P., Jun 2006. Multiscale white matter fiber tract coregistration: a new feature-based approach to align diffusion tensor data. *MRM* 55 (6), 1414–1423.
- Leemans, A., Sijbers, J., Verhoye, M., der Linden, A. V., Dyck, D. V., Apr 2005. Mathematical framework for simulating diffusion tensor mr neural fiber bundles. *Magn Reson Med* 53 (4), 944–953.
- Lenglet, C., 2006. Geometric and variational methods for Diffusion Tensor MRI processing. Ph.D. thesis, I.N.R.I.A.
- Lenglet, C., Campbell, J., Descoteaux, M., Haro, G., Savadjiev, P., Wassermann, D., Anwander, A., Deriche, R., Pike, G., Sapiro, G., Siddiqi, K., Thompson, P., 2009. Mathematical methods for diffusion mri processing. *Neuroimage* 45 (1), S111–S122.
- Lenglet, C., Deriche, R., Faugeras, O., 2004. Inferring white matter geometry from diffusion tensor MRI: Application to connectivity mapping. *Lecture Notes in Computer Science*, 127–140.
- Lenglet, C., Rousson, M., Deriche, R., June 2006a. DTI segmentation by statistical surface evolution. *Medical Imaging, IEEE Transactions on* 25 (6), 685–700.
- Lenglet, C., Rousson, M., Deriche, R., Faugeras, O., 2006b. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor mri processing. *J Math Imaging Vis* 25, 423–444.
- Levina, E., Bickel, P. J., 2004. Maximum likelihood estimation of intrinsic dimension. In: *NIPS*.
- Lin, F., Yu, C., Jiang, T., Li, K., Chan, P., 2007. Diffusion tensor tractography-based group mapping of the pyramidal tract in relapsing-remitting multiple sclerosis patients. *American Journal of Neuroradiology* 28 (2), 278.

- Lin, F., Yu, C., Jiang, T., Li, K., Li, X., Qin, W., Sun, H., Chan, P., 2006. Quantitative analysis along the pyramidal tract by length-normalized parameterization based on diffusion tensor tractography: application to patients with relapsing neuromyelitis optica. *NeuroImage* 33 (1), 154–160.
- Lindeberg, T., 1998. Edge Detection and Ridge Detection with Automatic Scale Selection. *IJCV* 30 (2), 117–156.
- Linden, D., 2002. Five hundred years of brain images.
- Lloyd, G., 1970. *Early Greek Science: Thales to Aristotle*. Chatto & Windus.
- Lori, N., Akbudak, E., Shimony, J., Cull, T., Snyder, A., Guillory, R., Conturo, T., 2002. Diffusion tensor fiber tracking of human brain connectivity: acquisition methods, reliability analysis and biological results. *NMR Biomed* 15, 493–515.
- Lovász, L., 1993. Random walks on graphs: a survey. In: *Paul Erdos is eighty*. Vol. 2. Combinatorics, János Bolyai Math So., Keszthely, pp. 353–397.
- MacKay, D. J. C., 1998. Introduction to gaussian processes. In: *Neural Networks and Machine Learning*. Vol. 168 of NATO ASI. Springer, pp. 133–165.
- Maddah, M., 2008. Quantitative analysis of cerebral white matter anatomy from diffusion mri. Ph.D. thesis, MIT.
- Maddah, M., Grimson, W. E. L., Warfield, S. K., Wells, W. M., 2008a. A unified framework for clustering and quantitative analysis of white matter fiber tracts. *MIA*, 191–202.
- Maddah, M., Wells III, W., Warfield, S., Westin, C., Grimson, W., 2007. Probabilistic clustering and quantitative analysis of white matter fiber tracts. *Int'l Conf. Information Processing in Medical Imaging*.
- Maddah, M., Zollei, L., Grimson, W. E. L., Westin, C.-F., Wells, W. M., 2008b. A mathematical framework for incorporating anatomical knowledge in DT-MRI analysis. In: *ISBI*. pp. 105–108.
- Malcolm, J., Rathi, Y., Tannenbaum, A., 2007. A graph cut approach to image segmentation in tensor space. In: *IEEE CVPR*.
- Malcolm, J. G., Michailovich, O., Bouix, S., Westin, C.-F., Shenton, M. E., Rathi, Y., 2010. A filtered approach to neural tractography using the Watson directional function. *Medical Image Analysis* 14, 58–69.
- Mann, H. B., Whitney, D. R., 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18, 50–60.

- Mansfield, P., 1977. Multi-planar image formation using nmr spin echoes. *Journal of Physics C: Solid State Physics* 10, L55–L58.
- McGraw, T., Vemuri, B., Yeziarski, R., Mareci, T., 2006. Segmentation of high angular resolution diffusion mri modeled as a field of von mises-fisher mixtures. *ECCV* 3953, 463–475.
- Meilă, M., 2006. The uniqueness of a good optimum for K-means. In: *Proceedings of the 23rd international conference on Machine learning*. ACM, pp. 632–640.
- Meilă, M., Shi, J., 2001. A random walks view of spectral segmentation. In: *AI and Statistics (AISTATS)*. pp. 92–96.
- Mesulam, M., 2005. Imaging connectivity in the human cerebral cortex: The next frontier? *Annals of Neurology* 57 (1), 5–7.
- Mohar, B., 1991. The Laplacian spectrum of graphs. *Graph Theory, Combinatorics, and Applications* 2, 871–898.
- Møller, M., Frandsen, J., Andersen, G., Gjedde, A., Vestergaard-Poulsen, P., Østergaard, L. R., 2007. Dynamic changes in corticospinal tracts after stroke detected by fibretracking. *British Medical Journal* 78 (6), 587.
- Monro, A. J., 1813. *Outlines of the anatomy of the human body, in its sound and diseased state*. Archibalde Constable and Co, Edinburgh.
- Mori, S., Crain, B. J., Chacko, V. P., Zijl, P. C. M. V., 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann Neur* 45 (2), 265–269.
- Mori, S., Wakana, S., Nagaе-Poetscher, L. M., van Zijl, P. C., 2005. *MRI Atlas of Human White Matter*. Elsevier.
- Moseley, M. E., Cohen, Y., Kucharczyk, J., Mintorovitch, J., Asgari, H. S., Wendland, M. F., Tsuruda, J., Norman, D., 1990. Diffusion-weighted mr imaging of anisotropic water diffusion in cat central nervous system. *Radiology* 176, 439–445.
- Nadler, B., Lafon, S., Coifman, R. R., Kevrekidis, I. G., 2006. Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators. In: *Weiss, Y., Schölkopf, B., Platt, J. (Eds.), Advances in Neural Information Processing Systems 18*. MIT Press, Cambridge, MA, pp. 955–962.
- Neji, R., Besbes, A., Komodakis, N., Deux, J., Maatouk, M., Rahmouni, A., Bassez, G., Fleury, G., Paragios, N., 2009. Clustering of the human skeletal muscle fibers using linear programming and angular hilbertian metrics. In: *Proceedings of the 21st International Conference on Information Processing in Medical Imaging*. Springer, p. 25.

- Ng, A., Jordan, M., Weiss, Y., 2002. On spectral clustering: Analysis and an algorithm. In: *Advances in Neural Information Processing Systems*. Vol. 14.
- Obeyesekere, G., 1977. The theory and practice of psychological medicine in the Ayurvedic tradition. *Culture, Medicine and Psychiatry* 1 (2), 155–181.
- O'Donnell, L. J., May 2006. Cerebral white matter analysis using diffusion imaging. Ph.D. thesis, Massachusetts Institute of Technology.
- O'Donnell, L. J., Westin, C.-F., nov 2007. Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE TMI* 26 (11), 1562–1575.
- O'Donnell, L. J., Westin, C.-F., Golby, A. J., 2007. Tract-based morphometry. In: [Ayache et al.: 2007], pp. 161–168.
- O'Donnell, L. J., Westin, C.-F., Golby, A. J., 2009. Tract-based morphometry for white matter group analysis. *NeuroImage* 45 (3), 832 – 844.
- Oh, J. S., Song, I. C., Lee, J. S., Kang, H., Park, K. S., Kang, E., Lee, D. S., Jul 2007. Tractography-guided statistics (tgis) in diffusion tensor imaging for the detection of gender difference of fiber integrity in the midsagittal and parasagittal corpora callosa. *Neuroimage* 36 (3), 606–616.
- Özarslan, E., Koay, C., Basser, P., 2009. Remarks on Q-space MR propagator in partially restricted, axially-symmetric, and isotropic environments. *Magnetic Resonance Imaging* 27 (6), 834–844.
- Paciorek, C., Schervish, M., 2006. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics (London, Ont.)* 17 (5), 483–506.
- Pagani, E., Filippi, M., Rocca, M., Horsfield, M., 2005. A method for obtaining tract-specific diffusion tensor MRI measurements in the presence of disease: application to patients with clinically isolated syndromes suggestive of multiple sclerosis. *Neuroimage* 26 (1), 258–265.
- Pagel, W., 1959. Medieval and Renaissance contributions to knowledge of the brain and its functions. In: *The History and Philosophy of Knowledge of the Brain and its Functions*.
- Pajevic, S., Pierpaoli, C., 1999. Color Schemes to Represent the Orientation of Anisotropic Tissues From Diffusion Tensor Data: Application to White Matter Fiber Tract Mapping in the Human Brain. *Magnetic Resonance in Medicine* 42, 526–540.
- Pampalk, E., Goebel, W., Widmer, G., 2003. Visualizing changes in the structure of data for exploratory feature selection. In: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM New York, NY, USA, pp. 157–166.

- Parker, G., Wheeler-Kingshott, C., Barker, G., 2002. Estimating distributed anatomical connectivity using fast marching methods and diffusion tensor imaging. *IEEE Transactions on Medical Imaging* 21 (5), 505–512.
- Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2 (6), 559–572.
- Peled, S., Friman, O., Jolesz, F., Westin, C.-F., nov 2006. Geometrically constrained two-tensor model for crossing tracts in DWI. *Magnetic Resonance Imaging* 24 (9), 1263–1270.
- Peled, S., Yeshurun, Y., 2001. Superresolution in MRI: Application to Human White Matter Fiber Tract Visualization by Diffusion Tensor Imaging. *Magnetic Resonance in Medicine* 45, 29–35.
- Pelleg, D., Moore, A., 2000. X-means: Extending K-means with efficient estimation of the number of clusters. In: *Proceedings of the Seventeenth International Conference on Machine Learning*. San Francisco, pp. 727–734.
- Perrin, M., Cointepas, Y., Cachia, A., Poupon, C., Thirion, B., Rivière, D., Cathier, P., El Kouby, V., Constantinesco, A., Le Bihan, D., Mangin, J.-F., 2008. Connectivity-based parcellation of the cortical mantle using Q-ball diffusion imaging. *International J Biomedical Imaging* 2008.
- Perrin, M., Poupon, C., Cointepas, Y., Rieul, B., Golestani, N., Pallier, C., 2005. Fiber Tracking in Q-ball Fields Using Regularized Particle Trajectories. In: *Information Processing in Medical Imaging*. Springer.
- Pierpaoli, C., Barnett, A., Pajevic, S., Chen, R., Penix, L., Virta, A., Basser, P., 2001. Water diffusion changes in Wallerian degeneration and their dependence on white matter architecture. *Neuroimage* 13 (6), 1174–1185.
- Pierpaoli, C., Jezzard, P., Basser, P., Barnett, A., Chiro, G. D., 1996. Diffusion Tensor MR imaging of human brain. *Radiology* 201, 637–648.
- Pothen, A., Simon, H. D., Liou, K.-P., 1990. Partitioning sparse matrices with eigenvectors of graphs. *SIAM Journal on Matrix Analysis and Applications* 11 (3), 430–452.
- Poupon, C., 1999. Détection des faisceaux de fibres de la substance blanche pour l'étude de la connectivité anatomique cérébrale. Ph.D. thesis, ECOLE NATIONALE SUPERIEURE DES TELECOMMUNICATIONS de PARIS.
- Poupon, C., Clark, C., Frouin, V., Regis, J., Bloch, I., Le Bihan, D., Mangin, J., 2000. Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles. *Neuroimage* 12 (2), 184–195.

- Poupon, C., Rieul, B., Kezele, I., Perrin, M., Poupon, F., Mangin, J., 2008. New diffusion phantoms dedicated to the study and validation of high-angular-resolution diffusion imaging (HARDI) models. *Magnetic resonance in Medicine* 60 (6), 1276.
- Purcell, E., Torrey, H., Pound, R., 1946. Resonance absorption by nuclear magnetic moments in a solid. *Physical Review* 69 (1-2), 37–38.
- Qazi, A. A., Radmanesh, A., O'Donnell, L., Kindlmann, G., Peled, S., Whalen, S., Westin, C.-F., Golby, A. J., 2009. Resolving crossings in the corticospinal tract by two-tensor streamline tractography: method and clinical assessment using fMRI. *Neuroimage* 47 (2), T98–T106.
- Ramón y Cajal, S., 1933. ¿neuronismo o rticularismo? las pruebas objetivas de la unidad anatómica de las células nerviosas. *Archivos de Neurobiología* 13, 570–646.
- Ramón y Cajal, S., 1995. *Histology of the nervous system of man and vertebrates*. Oxford University Press.
- Rasmussen, C., 2000. The infinite Gaussian mixture model. *Advances in neural information processing systems* 12, 554–560.
- Rasmussen, C. E., Williams, C. K. I., 2006. *Gaussian Processes for Machine Learning*. The MIT Press.
- Roebroek, A., Galuske, R., Formisano, E., Chiry, O., Bratzke, H., Ronen, I., Kim, D., Goebel, R., 2008. High-resolution diffusion tensor imaging and tractography of the human optic chiasm at 9.4 T. *Neuroimage* 39 (1), 157–168.
- Rosenberg, S., 1997. *The Laplacian on a Riemannian Manifold: An Introduction to Analysis on Manifolds*. Cambridge University Press.
- Roux, G., 1966. *Ancient Iraq*. Penguin.
- Roweis, S. T., Saul, L. K., 2000. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 290 (5500), 2323–2326.
- Sarlls, J., Pierpaoli, C., 2009. In vivo diffusion tensor imaging of the human optic chiasm at sub-millimeter resolution. *Neuroimage* 47 (4), 1244–1251.
- Sarton, G., 1993. *Ancient science through the golden age of Greece*. Dover Pubns.
- Savadjiev, P., Campbell, J., Descoteaux, M., Deriche, R., Pike, G., Siddiqi, K., 2008. Labeling of ambiguous subvoxel fibre bundle configurations in high angular resolution diffusion MRI. *NeuroImage* 41 (1), 58–68.
- Savadjiev, P., Kindlmann, G. L., Bouix, S., Shenton, M. E., Westin, C.-F., 2010. Local white matter geometry from diffusion tensor gradients. *NeuroImage* 49 (4), 3175 – 3186.

- Schaechter, J., Perdue, K., Wang, R., 2008. Structural damage to the corticospinal tract correlates with bilateral sensorimotor cortex reorganization in stroke patients. *Neuroimage* 39 (3), 1370–1382.
- Schmahmann, J., Pandya, D., 2007a. Cerebral white matter—historical evolution of facts and notions concerning the organization of the fiber pathways of the brain. *Journal of the History of the Neurosciences* 16 (3), 237–267.
- Schmahmann, J., Pandya, D., 2007b. The complex history of the fronto-occipital fasciculus. *Journal of the History of the Neurosciences* 16 (4), 362–377.
- Schmahmann, J., Pandya, D., 2009. *Fiber pathways of the brain*. Oxford Univ Pr.
- Schmahmann, J., Pandya, D., Wang, R., Dai, G., D’Arceuil, H., de Crespigny, A., Wedeen, V., 2007. Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* 130 (3), 630.
- Schmidt, E., 1908. Über die auflösung linearer gleichungen mit unendlich vielen unbekanntem. *Rend. Circ. Mat. Palermo* 25, 63–77.
- Scholkopf, B., Smola, A., Muller, K., 1999. Kernel principal component analysis. *Advances in Kernel Methods-Support Vector Learning*, 327–352.
- Scholkopf, B., Smola, A. J., Müller, K.-R., 1998. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation* 10 (5), 1299–1319.
- Schwartz, E., Cooper, E., Chin, C., Wehrli, S., Tessler, A., Hackney, D., 2005. Ex vivo evaluation of ADC values within spinal cord white matter tracts. *American Journal of Neuroradiology* 26 (2), 390.
- Scifo, P., Dell’Acqua, F., Rizzo, G., Gilardi, M., F., F., 2004. A dedicated phantom for diffusion tensor imaging studies. In: *ISMRM*.
- Seeger, M., 2004. Gaussian processes for machine learning. *International Journal of Neural Systems* 14 (2), 69–106.
- Sheikh, Y. A., Khan, E., Kanade, T., oct 2007. Mode-seeking by medoidshifts. In: *IEEE International Conference on Computer Vision*. No. 141.
- Shi, J., Malik, J., Aug. 2000. Normalized cuts and image segmentation. *IEEE PAMI* 22 (8), 888–905.
- Singer, C., 1956. Brain dissection before Vesalius. *Journal of the History of Medicine and Allied Sciences* 11 (3), 261.
- Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., Behrens, T. E. J., Jul 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 31 (4), 1487–1505.

- Song, S.-K., Sun, S.-W., Ramsbottom, M. J., Chang, C., Russell, J., Cross, A. H., 2002. Demyelination revealed through mri as increased radial (but unchanged axial) diffusion of water. *NeuroImage* 17 (3), 1429 – 1436.
- Song, S.-K., Yoshino, J., Le, T. Q., Lin, S.-J., Sun, S.-W., Cross, A. H., Armstrong, R. C., 2005. Demyelination increases radial diffusivity in corpus callosum of mouse brain. *NeuroImage* 26 (1), 132 – 140.
- Sosa, J., Valderrama-Isturbe, P., 2001. Psychology in Latin America: Historical reflections and perspectives. *International Journal of Psychology* 36 (6), 384–394.
- Souvenir, R., Pless, R., 2005. Manifold Clustering. In: *IEEE ICCV*. pp. 648—653.
- Stein, M. L., 1999. *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. Springer.
- Steinhaus, H., 1956. Sur la division des corp materiels en parties. *Bull. Acad. Polon. Sci* 1, 801–804.
- Stejskal, E., Tanner, J., 1965. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J Chemical Physics* 42, 288—292.
- Stieltjes, B., Kaufmann, W. E., van Zijl, P. C. M., Fredericksen, K., Pearlson, G. D., Solaiyappan, M., Mori, S., 2001. Diffusion tensor imaging and axonal tracking in the human brainstem. *NeuroImage* 14 (3), 723 – 735.
- Stroman, P., Kornelsen, J., Lawrence, J., 2005. An improved method for spinal functional MRI with large volume coverage of the spinal cord. *J. Mag Res Imaging* 21 (5).
- Swift, J., 1726. *Gulliver's travels*. Blackwell.
- Talairach, J., Tournoux, P., January 1988. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System : An Approach to Cerebral Imaging*. Thieme Medical Publishers.
- Tanner, J., Stejskal, E., 1968. Restricted Self-Diffusion of Protons in Colloidal Systems by the Pulsed-Gradient, Spin-Echo Method. *The Journal of Chemical Physics* 49, 1768.
- Tascioglu, A., Tascioglu, A., 2005. Ventricular anatomy: illustrations and concepts from antiquity to Renaissance. *Neuroanatomy* 4, 57–63.
- Taylor, D., Bushell, M., 1985. The spatial mapping of translational diffusion coefficients by the NMR imaging technique. *Physics in medicine and biology* 30, 345–349.

- Teh, Y., Jordan, M., Beal, M., Blei, D., 2006. Hierarchical dirichlet processes. *Journal of the American Statistical Association* 101 (476), 1566–1581.
- Tenenbaum, J., Silva, V., Langford, J., 2000. A Global Geometric Framework for Non-linear Dimensionality Reduction. *Science* 290 (5500), 2319.
- Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 63 (2), 411–423.
- Tishby, N., Slonim, N., 2000. Data clustering by markovian relaxation and the information bottleneck method. In: *Neural Information Processing Systems*.
- Todd, R., 1845. *The Descriptive and Physiological Anatomy of the Brain, Spinal Cord and Ganglions*. Sherwood, Gilbert and Piper.
- Tofts, P., Lloyd, D., Clark, C., Barker, G., Parker, G., McConville, P., Baldock, C., Pope, J., 2000. Test liquids for quantitative MRI measurements of self-diffusion coefficient in vivo. *Magnetic Resonance in Medicine* 43 (3), 368–374.
- Torgerson, W., 1952. Multidimensional scaling: I. Theory and method. *Psychometrika* 17 (4), 401–419.
- Tournier, J. D., Calamante, F., Gadian, D. G., Connelly, A., 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* 23, 1176–1185.
- Tristán-Vega, A., Aja-Fernández, S., 2009. DWI filtering using joint information for DTI and HARDI. *Medical Image Analysis*.
- Tristán-Vega, A., Westin, C., Aja-Fernández, S., 2009. Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging. *NeuroImage* 47 (2), 638–650.
- Tuch, D., Reese, T., Wiegell, M., Makris, N., Belliveau, J., Wedeen, V., 2002. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine* 48 (4), 577–582.
- Tuch, D., Wisco, J., Khachaturian, M., Ekstrom, L., Koö tter, R., Vanduffel, W., 2005. Q-ball imaging of macaque white matter architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1457), 869–879.
- Tuch, D. S., 2002. Diffusion MRI of complex tissue structure. Ph.D. thesis, massachusetts institute of technology.
- Tuch, D. S., Dec 2004. Q-ball imaging. *Magn Reson Med* 52 (6), 1358–1372.

- Veltkamp, R., 2001. Shape matching: Similarity measures and algorithms. In: *Shape Modeling and Applications*, SMI. pp. 188–197.
- Vesalius, A., 1563. *De humani corporis fabrica*. Venecia. Giunti.
- von Luxburg, U., Bousquet, O., Belkin., M., 2005. Limits of spectral clustering. In: Saul, L. K., Weiss, Y., Bottou, L. (Eds.), *Advances in Neural Information Processing Systems (NIPS)*. MIT Press.
- Von Staden, H., 1989. *Herophilus: the art of medicine in early Alexandria: edition, translation, and essays*. Cambridge Univ Pr.
- Wahba, G., 1990. *Spline Models for Observational Data*. Soc Ind Math.
- Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., Hua, K., Zhang, J., Jiang, H., Dubey, P., Blitz, A., van Zijl, P., Mori, S., 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. *NeuroImage* 36 (3), 630 – 644.
- Wakana, S., Jiang, H., Nagae-Poetscher, L., van Zijl, P., Mori, S., 2004. Fiber Tract-based Atlas of Human White Matter Anatomy. *Radiology* 230, 77–87.
- Wang, Z., Vemuri, B., 2004. Tensor Field Segmentation Using Region Based Active Contour Model. In: *ECCV*. Springer.
- Wang, Z., Vemuri, B., 2005. DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE transactions on medical imaging* 24 (10), 1267–1277.
- Wang, Z., Vemuri, B., Chen, Y., Mareci, T., 2004. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. *Medical Imaging, IEEE Transactions on* 23 (8), 930–939.
- Wassermann, D., Bloy, L., Kanterakis, E., Verma, R., Deriche, R., 2010a. Unsupervised white matter fiber clustering and tract probability map generation: Applications of a gaussian process framework for white matter fibers. *NeuroImage* 51 (1), 228–241.
- Wassermann, D., Bloy, L., Verma, R., Deriche, R., 2009. A Gaussian Process based framework for white matter fiber tracts and bundles, applications to fiber clustering. In: *MICCAI Workshops*. London, UK.
- Wassermann, D., Cohen-Adad, J., Lehericy, S., Benali, H., Rossignol, S., Deriche, R., 2010b. Fully automated straightening of the spinal cord using fiber tractography. In: *18th. International Symposium on Magnetic Resonance in Medicine*.
- Wassermann, D., Cohen-Adad, J., Lehericy, S., Benali, H., Rossignol, S., Deriche, R., 2010c. Straightening the spinal cord using fiber tractography. In: *International Symposium on Biomedical Imaging*.

- Wassermann, D., Deriche, R., 2008. Simultaneous manifold learning and clustering: Grouping white matter fiber tracts using a volumetric white matter atlas. In: MICCAI Workshops.
- Wassermann, D., Descoteaux, M., Deriche, R., 2007. Diffusion maps segmentation of magnetic resonance Q-Ball imaging. In: IEEE Workshop on Mathematical Methods in Biological Imaging.
- Wassermann, D., Descoteaux, M., Deriche, R., 2008. Diffusion maps clustering for magnetic resonance Q-Ball imaging segmentation. *International J Biomedical Imaging* 2008.
- Wassermann, D., Kanterakis, E., Gur, R. C., Deriche, R., Verma, R., 2010d. Diffusion-based population statistics using tract probability maps. In: Jiang, T. (Ed.), MICCAI. No. 6361 in LNCS. Springer Berlin / Heidelberg, pp. 630–638.
- Websey, G., Moseley, M., Ehman, R., 1984a. Translational Molecular Self-Diffusion in Magnetic Resonance Imaging: I. Effects on Observed Spin-Spin Relaxation. *Investigative Radiology* 19 (6), 484.
- Websey, G., Moseley, M., Ehman, R., 1984b. Translational Molecular Self-Diffusion in Magnetic Resonance Imaging: II. Measurement of the Self-Diffusion Coefficient. *Investigative Radiology* 19 (6), 491.
- Wedeen, V., Wang, R., Schmahmann, J., Benner, T., Tseng, W., Dai, G., Pandya, D., Hagmann, P., D'Arceuil, H., de Crespigny, A., 2008. Diffusion spectrum magnetic resonance imaging (ds) tractography of crossing fibers. *NeuroImage*.
- Wei, Y., Cheng, C., 1991. Ratio cut partitioning for hierarchical designs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 10 (7), 911–921.
- Weldeselassie, Y. T., Hamarneh, G., 2007. Dt-mri segmentation using graph cuts. In: SPIE Conference in Medical Imaging: Image Processing,.
- Wendland, H., 1995. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics* 4 (1), 389–396.
- Westin, C.-F., Maier, S. E., Mamata, H., Nabavi, A., Jolesz, F. A., Kikinis, R., Jun 2002. Processing and visualization for diffusion tensor mri. *Med Image Anal* 6 (2), 93–108.
- Wiegell, M., Larsson, H., Wedeen, V., 1999. Diffusion tensor MRI of the thalamus: differentiation of nuclei by their projections. *Proceedings of the International Society for Magnetic Resonance in Medicine* 934.

- Wiegell, M., Tuch, D., Larsson, H., Wedeen, V., 2003. Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging. *Neuroimage* 19 (2), 391–401.
- Wiegell, M. R., Larsson, H. B., Wedeen, V. J., Dec 2000. Fiber crossing in human brain depicted with diffusion tensor mr imaging. *Radiology* 217 (3), 897–903.
- Wilkins, R. H., March 1964. Neurosurgical classic-XVII: Edwin Smith surgical papyrus. *Journal of Neurosurgery* 21 (3), 240–244.
- Williams, O., Fitzgibbon, A., 2007. Gaussian process implicit surfaces. In: *Gaussian Proc. in Practice*.
- Williams, T., Gluhbegovic, N., Jew, J., 1997. *The human brain: dissections of the real brain*. Vitruval Hospital, University of Iowa.
- Worsley, K., Marrett, S., Neelin, P., Vandal, A., Friston, K., Evans, A., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping* 4 (1), 58–73.
- Yanasak, N., Allison, J., 2006. Use of capillaries in the construction of an MRI phantom for the assessment of diffusion tensor imaging: demonstration of performance. *Magnetic Resonance Imaging* 24 (10), 1349–1361.
- Yang, J., Shen, D., Davatzikos, C., Verma, R., 2008. Diffusion Tensor Image Registration Using Tensor Geometry and Orientation Features. In: Metaxas, D. (Ed.), *MICCAI*. No. 5241 in LNCS. Springer Berlin / Heidelberg, pp. 905–913.
- Yu, S., Shi, J., 2003. Multiclass spectral clustering. In: *International conference on computer vision*. Vol. 5. Citeseer.
- Yushkevich, P. A., Zhang, H., Simon, T. J., Gee, J. C., 2008. Structure-specific statistical mapping of white matter tracts. *NeuroImage* 41 (2), 448 – 461.
- Zelnik-Manor, L., Perona, P., 2004. Self-tuning spectral clustering. *Advances in Neural Information Processing Systems* 17, 1601–1608.
- Zhukov, L., Museth, K., Breen, D., Whitaker, R., Barr, A., 2003. Level set modeling and segmentation of DT-MRI brain data. *J Electronic Imaging* 12 (1), 125–133.
- Ziyan, U., Tuch, D., Westin, C.-F., 2006. Segmentation of thalamic nuclei from dti using spectral clustering. In: *MICCAI*. LNCS. Springer/Verlag, pp. 807–814.