

**COMPUTATIONAL STATISTICS FOR HUMAN BRAIN
DIFFUSION TENSOR IMAGE ANALYSIS**

SAFA ELSHEIKH

A thesis submitted in partial fulfilment of the
requirements of the University of Brighton
for the degree of Doctor of Philosophy

November 2018

COMPUTATIONAL STATISTICS FOR HUMAN BRAIN
DIFFUSION TENSOR IMAGE ANALYSIS

SAFA ELSHEIKH

PhD

2018

Contents

Acknowledgments	vi
Declarations	vii
Abstract	viii
Abbreviations	x
Chapter 1 Introduction	1
1.1 Problem Overview	1
1.2 Motivation	4
1.3 Research Questions	5
1.4 Research Methodology	6
1.5 Contribution to Knowledge	8
1.6 Thesis Organisation	9
1.7 Publications	9
Chapter 2 Background and Related Work	11
2.1 Magnetic Resonance Imaging	11
2.2 Diffusion Weighted Imaging	14
2.3 Diffusion Tensor Imaging	15
2.3.1 Basic Concepts	16
2.3.2 Estimation of Diffusion Tensor	17
2.3.3 Eigensystem and Ellipsoid	20

2.3.4	Diffusion Anisotropy Measures	21
2.3.5	The Views of the Brain	24
2.3.6	DTI Processing	25
2.3.7	DTI Applications	29
Chapter 3	Cluster Analysis of DTI	32
3.1	Introduction	33
3.2	Background	35
3.2.1	Euclidean and Non-Euclidean Methods	35
3.2.2	DTI Processing Using Non-Euclidean Methods	42
3.2.3	Segmentation of the CC Using Non-Euclidean Methods	43
3.2.4	Processes on the Covariance Matrices	44
3.3	Cluster Analysis	45
3.3.1	Hierarchical Clustering	45
3.3.2	Description of K -means	50
3.3.3	Hartigan's Method	51
3.4	Hartigan's method with f -mean	55
3.4.1	F -mean	55
3.4.2	Hartigan's Method with the f -mean	56
3.4.3	Special Cases	62
3.5	Applications	63
3.5.1	Segmentation using Hierarchical Clustering	64
3.5.2	Two Dimensional Segmentation Using Hartigan's Method . . .	66
3.5.3	Three Dimensional Segmentation Using Hartigan's Method . .	71
3.6	Discussion and Conclusions	75
Chapter 4	Modeling Diffusion Directions of the CC	78
4.1	Introduction	79
4.2	Background and Related Work	80
4.2.1	Multiple Sclerosis	80
4.2.2	DTI measures in MS	82

4.3	Von Mises-Fisher Distribution	83
4.3.1	Maximum Likelihood Estimators	85
4.3.2	Mixture of vmf	87
4.4	Modeling Diffusion Directions of the CC	89
4.4.1	Simulation Study	90
4.4.2	Brain Images	91
4.4.3	Methodology	91
4.5	Results	92
4.6	Conclusion and Discussion	99
Chapter 5	Fiber Tracking of the CC	102
5.1	Introduction and Related Work	103
5.2	Deterministic White Matter Tractography	106
5.3	Probabilistic Tractography	108
5.3.1	Probability Density Function (pdf)	110
5.3.2	Probabilistic Local Tracking Methods	111
5.4	Stopping Criteria	111
5.5	Application	112
5.5.1	Evaluation of Stopping Criteria	113
5.5.2	Deterministic and Probabilistic Tracking	118
5.6	Conclusions and Discussions	124
Chapter 6	Conclusions and Future Work	128
6.1	Summary and Conclusions	128
6.2	Discussions and Future Work	131
6.2.1	Evaluation of the Segmentation of the CC	131
6.2.2	Classification of Brain Tissue	132
6.2.3	Clustering the Brain Lobes	132
6.2.4	Diffusion Directions	133

Appendix A Appendices	135
A.1 Euclidean and non-Euclidean Distances	135
A.2 Euclidean and non-Euclidean Means	137
A.3 Particular Cases of the f -mean	138
A.4 Fiber tracking	139

Acknowledgments

I would like to express my deep gratitude to my supervisors Dr. Andrew Fish, Dr. Roma Chakrabarti and Dr. Diwei Zhou for the great support, professional guidance and the excellent advice provided during the course of my PhD studies. I am very happy I have such great supervisors. I would like to thank Professor Mara Cercignani, Dr. Samira Bouyagoub and Dr. Nicholas Dowell for their valuable advice and assistance. My special thanks to Professor Mara Cercignani for providing the brain data.

Thanks to all of my friends and colleagues in UK and Sudan. I would also like to acknowledge the continuous support and encouragement provided by all of my family. I would like to express my very great appreciation to Dr. Diwei Zhou for the continuous encouragement, the great advice and for providing the required software and data. Special thanks to Dr. Andrew Fish for his great, fast and detailed feedback and valuable suggestions. His willingness to help and give his time so generously has been very much appreciated.

The work was partially supported by University of Khartoum, Sudan and the University of Brighton.

Declarations

I declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to this or any other university for a degree, and does not incorporate any material already submitted for a degree.

Abstract

Diffusion tensor imaging (DTI) is an advanced technique of magnetic resonance imaging (MRI) which is able to measure the diffusion of the water inside the brain tissues. Developing statistical methods for accurate grouping and modeling data objects with complex nature, such as the space of diffusion tensors, is needed to improve disease diagnosis and surgical planning. In this thesis, new statistical methodologies for DTI of human brain are developed.

The corpus callosum (CC) is a great fiber bundle in the white matter of the brain. Accurate segmentation of the CC is an important aspect of clinical medicine and is used in the diagnosis of various brain disorders. An accurate automated method for two and three-dimensional segmentation of the CC using DTI is developed. Hartigan's K -means, an accurate K -means algorithm, is generalized for use with f -mean metrics (e.g. Cholesky, root Euclidean and log Euclidean). Then the generalized algorithm is used to provide a segmentation of the CC. The segmentation results using different metrics are evaluated to determine which metrics lead to the most accurate segmentations.

The von Mises-Fisher distribution (vmf) is a probability distribution for modeling directional data on the unit hypersphere. Multiple sclerosis (MS) is a neuro-inflammatory disease that affects the brain and spinal cord and it is considered the most common neurological disease that cause disabilities in young adults. We modeled the diffusion directions of the CC as a mixture of vmf distributions for MS and healthy subjects. Higher diffusion concentration around the mean directions and smaller sum of angles between the mean directions are observed on the normal-

appearing CC of the MS as compared to the healthy subjects. An individual-based curvature threshold for stopping criteria of fiber tracking in the CC is proposed.

Fiber tracking is an important tool for visualizing white matter pathways and detecting brain abnormalities. Deterministic fiber tracking is highly affected by the noise in diffusion weighted imaging (DWI) and abnormalities in the brain which cause errors in the propagation of the track. The proposed stopping criteria, which is based on individual's curvature thresholds, improves fiber tracking by terminating the fiber to prevent the deviation out of the original pathway. Quantitative measures (fractional anisotropy (FA), mean diffusivity (MD) and length of fibers) are compared in the healthy and MS subjects using three fiber tracking techniques (FACT, Bayesian and Wild Bootstrap). Significant differences in length of fibers in the healthy and the MS subjects are obtained using the three tracking methods.

Overall, the main contributions of the thesis is the development of new statistical methods for accurate clustering and modeling of data using DTI taking into account the complex nature of the data.

Abbreviations

ACM - Anatomical connectivity mapping

ADC - apparent diffusion coefficient

autism- autistic spectrum disorder

BIC - Bayesian information criterion

CC - Corpus Callosum

CLLS - constrained linear least square estimation

CNLS - constrained non-linear least square estimation

CSF - cerebrospinal fluid

DTI - Diffusion Tensor Imaging

DWI - Diffusion Weighted Imaging

EPI - echo planar imaging

EDSS - Expanded Disability Status Scale

FA - Fractional Anisotropy

FACT - Fiber Assignment by Continuous Tracking

HARDI - High Angular Resolution Diffusion Imaging

LLS - linear least square estimation

MRI - Magnetic Resonance Imaging

MD - Mean Diffusivity

MS - Multiple sclerosis

NLS - non-linear least square estimation

odf - orientation distribution function

PCA - Principal Component Analysis

pdf - Probability density function
PGA - Principal Geodesic Analysis
RA - Relative Anisotropy
RF - radio frequency
RMSE - root mean square error
ROI - region of interest
SH - spherical harmonic
SNR - signal to noise ratio
TEND - tensor deflection
VR - Volume Ratio
vmf - von Mises-Fisher
WCSS - within cluster sum of squares

Chapter 1

Introduction

The aim of the thesis is to develop computational statistics to accurately cluster and model complex data using Diffusion Tensor Imaging (DTI) which are useful in many clinical applications such as presurgical planning, intraoperative and detection of various brain disorders.

In this chapter, an overview of the thesis's research problems, motivations and research questions are provided in Sections 1.1, 1.2 and 1.3. Research Methodology and contributions to knowledge are presented in Section 1.4 and 1.5. The organisation of the thesis and the list of published papers during the research period are provided in Sections 1.6 and 1.7.

1.1 Problem Overview

Cluster analysis is important for statistical data analysis. There are a wide range of clustering applications (e.g. image processing, machine learning, bioinformatics and information retrieval). Most cluster analysis algorithms are designed for data with simple structure. Data with complex structure such as the space of positive semi definite symmetric matrices arise in many applications (e.g. image analysis, longitudinal data analysis and network data analysis). Clustering the data itself (i.e. matrices) is expected to be more accurate than clustering the points obtained by performing dimensionality reduction on the data, as no information will be lost from the

data. For the space of positive semi-definite symmetric matrices, non-Euclidean metrics (log Euclidean, Riemannian, Cholesky, Procrustes, root Euclidean and power Euclidean) are suitable alternatives to the Euclidean metric (Dryden, Koloydenko, & Zhou, 2009). Clustering methods based on Riemannian metric have been proposed (e.g. Rattray, 2000; Goh & Vidal, 2008c; Goh & Vidal, 2008a). However the proposed methods either use points or perform a dimensionality reduction to cluster symmetric positive semi-definite matrices. A kernel K -means for positive definite covariance matrices using Riemannian, log Euclidean, Cholesky and power Euclidean metrics has been explored recently (Jayasumana, Hartley, Salzmann, Li, & Harandi, 2013). However, they did not use the Procrustes metric. An accurate version of K -means algorithm has been proposed by (Hartigan, 1975). Hartigan’s method provides results with tighter clusters than the Voronoi diagram (Telgarsky & Vattani, 2010) and hence the method converges to a solution with smaller within cluster sum of squares (WCSS) as compared with Lloyd’s method (Lloyd, 1982). The Lloyd method considers how close an object is to the centroid of the cluster and each object is assigned to its closest centroid. The result of Hartigan’s method guarantees that no movement of an object to any cluster will reduce the WCSS. Using Hartigan’s method to cluster the space of positive semi-definite symmetric matrices has not been explored in the literature.

The corpus callosum (CC) is a great fiber bundle in the white matter of the brain. A change in size and shape of the CC can be used as an indicator of a brain disorder (Duara et al., 1991; Hardan, Minshew, & Keshavan, 2000; Downhill et al., 2000; Walterfang et al., 2011; Bachman, Lee, Sidtis, & Ardekani, 2014). An accurate segmentation of the CC is needed in clinical applications (e.g. disease prognosis and surgical planning). The segmentation of the 2D CC using magnetic resonance imaging (MRI) and DTI has been studied in the literature. However, the 3D segmentation of the CC has been studied by few researchers using MRI (El-Baz et al., 2011) and DTI (Lenglet, Rousson, & Deriche, 2006; Nazem-Zadeh et al., 2012). The segmentation of the CC using the Riemannian metric is superior to the seg-

mentation using Euclidean metric (Lenglet et al., 2006). However, the Riemannian mean is computationally expensive as it needs a numerical method (e.g. a gradient descent algorithm) to compute (Pennec, Fillard, & Ayache, 2006). Due to high volume of brain images, the efficiency is important especially for the 3D segmentation. The log Euclidean metric yields similar results to the Riemannian metric and the log Euclidean mean is easier and faster to compute without a numerical method (Arsigny, Fillard, Pennec, & Ayache, 2006). Prior to the work in this thesis, there was no automatic 3D segmentation method of the CC using the log Euclidean metric.

The von Mises-Fisher distribution (vmf) is one of the most basic probability distributions for modeling directional data on the unit hypersphere (Mardia & Jupp, 2009). In Banerjee, Dhillon, Ghosh, and Sra (2005), the use of a finite mixture of vmf distributions to cluster directional data on the unit sphere has been proposed. The vmf distribution has been used in the literature to model the orientation distribution function (odf) for each voxel in the brain (e.g. McGraw, Vemuri, Yeziarski, & Mareci, 2006; Bhalerao & Westin, 2007; Kumar, Barmapoutis, Vemuri, Carney, & Mareci, 2008; Reynolds, Nir, Jahanshad, Prasad, & Thompson, 2014). However, modeling the diffusion directions for a region (not just a voxel) in the brain (such as the CC) using directional distributions is beneficial in recognizing the main diffusion direction and how the diffusion is concentrated on that region. This might help in detecting brain abnormalities. Such modeling also helps in setting an individual threshold for the curvature as stopping criteria for fiber tracking in that region (see next paragraph). Modeling diffusion directions for a region in the brain using directional distributions, including vmf distribution, has not been investigated in the literature.

Fiber tracking is important for visualizing the white matter tracts and understanding the connectivity between regions in the brain. Fiber tracking represents a useful tool for detecting brain abnormalities, as the white matter tracts are affected by many brain disorders. Many factors such as starting point and stopping

criteria lead to bias in the results of fiber tracking. Controlling these factors leads to reduction of the bias in the results (Girard, Whittingstall, Deriche, & Descoteaux, 2014). Common stopping criteria are anisotropy threshold and curvature threshold. Binary masks have been used as stopping criteria, however, they lead to poor results (Smith, Tournier, Calamante, & Connelly, 2012; Girard et al., 2014). A fixed threshold for stopping criteria can lead to bias in the results of fiber tracking Brecheisen, Vilanova, Platel, and ter Haar Romeny (2009). An individual-based anisotropy threshold for tracing fibers of the CC in MS is used by Brecheisen et al. (2009). However, individual-based curvature threshold has not been studied in the literature.

1.2 Motivation

Developing new statistical methods for accurate grouping and modeling data using DTI is important for accurate segmentation of a region in the brain, analysis of diffusion data and to improve fiber tracking results which are all useful and needed for accurate diagnosis of various brain disorders and are also important in clinical applications. These motivations are detailed as follows:

- Data with complex structure such as the space of positive semi-definite symmetric matrices arises in many applications (e.g. image analysis using diffusion tensor field, longitudinal data analysis). Clustering this data itself is expected to provide more accurate results, as we make use of all the information in the data, whereas dimensionality reduction destroy some of the original information in the data.
- Change in size or shape of the CC is an indicator of a brain abnormality (e.g. Hardan et al., 2000; Bachman et al., 2014). Hence an accurate 2D and 3D segmentation method for the CC is incredibly important to improve diagnosis

of such diseases and for surgical planning.

- Change in diffusion pattern in the brain is an indicator of a brain disorder (Filippi, Cercignani, Inglese, Horsfield, & Comi, 2001; S.-K. Song et al., 2005; Makki, Chugani, Janisse, & Chugani, 2007). Hence analysing diffusion directions in the brain is useful in detection of such brain abnormalities.
- Fiber tracking is important as a non-invasive technique for visualizing white matter pathways in the brain which is useful for presurgical planning and intra-operative (Farquharson et al., 2013) and for detection of many abnormalities in the brain (P. J. Basser, Pajevic, Pierpaoli, Duda, & Aldroubi, 2000; Pagani, Filippi, Rocca, & Horsfield, 2005; Dubois, Hertz-Pannier, Dehaene-Lambertz, Cointepas, & Le Bihan, 2006). However, fiber tracking especially deterministic tracking is affected by noise and abnormalities in the brain. Modeling diffusion directions provides individual-based information which helps to improve fiber tracking in the brain by preventing fibers from deviation into incorrect pathways.

1.3 Research Questions

The main aim of the thesis is to develop new statistical methodologies for accurate grouping and modeling data using DTI. The main research questions are listed as follows:

- How can Hartigan’s method be generalized for the use with non-Euclidean methods to cluster diffusion tensors?
- How to develop a 2D and 3D segmentation method of the CC using the generalized Hartigan’s method? What are the most accurate segmentation results?

- How to model diffusion directions in a ROI in the human brain (i.e. the CC) using vmf distribution?
- How can modeling of the CC be used to develop a stopping criteria of fiber tracking of the CC?

1.4 Research Methodology

This section gives a general description of the methods used in the thesis. The statistical analysis throughout the thesis is performed using Minitab statistical software (Minitab 17, Inc., State College, PA) (Minitab, 2010). The methodology is divided into three steps as follows:

Firstly, Hartigan’s method is generalized for use with f -mean metrics to cluster diffusion tensors (and covariance matrices generally). The generalized Hartigan’s method is applied to segment the CC of a healthy brain. The healthy brain image was provided by the Academic Radiology Department of Queens Medical Center. Diffusion tensors are estimated using linear estimation method (P. Basser, Mattiello, & LeBihan, 1994). The number of clusters for segmentation the CC are evaluated using the silhouette method and the variance of the CC. A ROI that includes the CC is selected for each of the 2D and 3D segmentation. Euclidean and non-Euclidean methods are applied to provide 2D and 3D segmentations of the CC. The 2D segmentation is performed on the midsagittal plane. The 2D and 3D segmentation results obtained by using the Euclidean and non-Euclidean methods are evaluated. The Lloyd’s algorithm is used for the initialization of clustering to reduce the time of clustering. The focus here on Hartigan’s method, however the agglomerative hierarchical clustering (single-link) is also applied to segment the 2D CC. The implementations were performed using MATLAB (version 7.14.0.739 (R2012a), The MathWorks, Inc., Natick, Massachusetts, United States). The FA and size of the tensors in the 3D CC of the different methods are compared. In the second step, the proposed segmentation using the Euclidean method is applied to provide

a 3D segmentation of a group of healthy and MS subjects to model the diffusion directions of the CC (the Euclidean method is chosen as it provides the fastest 3D segmentation but any 3D segmentation method could be used).

Secondly, the diffusion directions of the CC are modeled as a mixture of vmf distributions for MS and healthy subjects. The data was collected at the Neuroimaging Laboratory, Santa Lucia Foundation in Rome (Italy). The study was approved by Ethics Committee of Santa Lucia Foundation. Written informed consent was obtained from all subjects before entering the study. The data consists of four healthy (two males, two females) and nine secondary progressive MS (three males, six females) subjects. Diffusion tensors are fitted linearly. The three dimensional segmentation of the CC is obtained, for the MS and healthy subjects, using the Euclidean method. The principal eigenvectors of the diffusion tensors are parallel to the diffusion directions. Hence, the data used for modeling are the principal eigenvectors of the diffusion tensors. The number of the mixture components is evaluated. The parameters of the model as well as FA and variance of FA are used to compare between the healthy and the MS subjects. The results are obtained using R programming language (R Core Team, 2013)-Package Directional. The maximum angle between the mean directions is proposed as a stopping criteria (individual-based curvature threshold) for fiber tracking in the CC (third step).

Thirdly, Fiber tracking of the CC is performed using the same data described in the second step. The Camino toolkit (Cook et al., 2006) is used for obtaining and analyzing fiber tracking results. For all the subjects, seed points are selected from the midsagittal slice on the middle of the CC using the software ITK-SNAP (Yushkevich et al., 2006). For visualization of tracking results, ParaView (Ayachit, 2015) software is used. The proposed stopping criteria is evaluated. Quantitative measures (i.e. FA, mean diffusivity (MD) and length of fibers) obtained from three fiber tracking methods: FACT, Bayesian (Friman, Farneback, & Westin, 2006) and Wild Bootstrap (Whitcher, Tuch, Wisco, Sorensen, & Wang, 2008) are compared

using the data of the healthy and MS subjects.

1.5 Contribution to Knowledge

In this thesis, we develop new statistical methodologies for accurate clustering and modeling data using DTI. Accurate clustering of DTI is useful in providing a more improved segmentation of a ROI in the brain as compared with segmentation using FA. Modeling data using DTI is useful in detecting abnormalities and to improve fiber tracking in the brain. The main contribution can be summarized as follows:

- Hartigan's method is generalized for the use with f -means metrics. It is shown that Hartigan's method cannot be generalized to use for metrics that require a numerical method for the computation of the mean (such as Riemannian and Procrustes). By adding a conditional statement to the generalized Hartigan's algorithm, the Riemannian and Procrustes metrics can be used only if the WCSS is always decreasing. However, there is no guarantee that the clustering converges to a local minima using those metrics.
- An automated segmentation method of the CC using Hartigan's method is developed. A suitable number of clusters for the segmentation of the CC is obtained. The segmentation of the CC using Euclidean and non-Euclidean methods are compared.
- Diffusion directions of the CC are modeled as a mixture of vmf distributions. It is shown that there are at least three main diffusion directions in the CC. Parameters of the mixture model are compared in healthy and MS subjects.
- Stopping criteria of fiber tracking of the CC based on curvature threshold is developed and evaluated.
- For fiber tracking in the CC, it is shown that any of the quantitative measures (FA, MD and length of fibers) obtained using FACT, Bayesian and Wild Bootstrap methods are not significantly different.

1.6 Thesis Organisation

In this chapter, we provided a general introduction to the thesis. The rest of the thesis's chapters are organized as follows:

Chapter 2: Background and Related Work

A general background of the thesis is provided in this chapter. This includes an overview of MRI, Diffusion weighted imaging (DWI) and DTI.

Chapter 3: Cluster Analysis of DTI

In this chapter, an automated method for segmentation of the CC using the generalized Hartigan's method is developed and the results of segmentation of the CC in a healthy brain are evaluated. The hierarchical clustering is also applied to provide a 2D segmentation of the CC.

Chapter 4: Modeling Diffusion Directions of the CC

The CC is segmented using the method proposed in Chapter 3, then the diffusion directions of the CC are modeled as a mixture of vmf distributions for multiple sclerosis (MS) and healthy subjects. Stopping criteria for fiber tracking in the CC is proposed.

Chapter 5: Fiber Tracking of the CC

In this chapter, the proposed stopping criteria for fiber tracking of the CC (Chapter 4) is evaluated. The results of three fiber tracking algorithms (FACT, Bayesian and Wild Bootstrap) are compared quantitatively. Fiber tracking of the MS and the healthy subjects are compared.

Chapter 6: Conclusions and Future Work

This chapter summarizes the results and the main findings. It also provides some discussions and directions for future work.

1.7 Publications

During PhD studies, two papers are published:

- Elsheikh, S., Fish, A., Chakrabarti, R., Zhou, D., Cercignani, M. (2017, July).

Modeling Diffusion Directions of Corpus Callosum. In Annual Conference on Medical Image Understanding and Analysis (pp. 518-526). Springer, Cham. This paper summarizes some of the work in Chapter 4.

- Elsheikh, S., Fish, A., Chakrabarti, R., Zhou, D. (2016). Cluster analysis of Diffusion Tensor fields with application to the segmentation of the Corpus Callosum. *Procedia Computer Science*, 90, 15-21. This paper summarize some of the work in Chapter 3.

Chapter 2

Background and Related Work

This chapter aims to provide a background of the Diffusion Tensor Imaging (DTI). Magnetic Resonance Imaging (MRI) and Diffusion Weighted Imaging (DWI) are reviewed first. Then basic concepts and estimation methods for diffusion tensors are provided. The eigensystem and ellipsoid representation of diffusion tensor are reviewed. Diffusion anisotropy measures are discussed. In addition, the possible views of the brain are explored. The processes on DTI and the possible applications of DTI are provided.

In this chapter, an introduction to MRI and DWI are provided in Sections 2.1 and 2.2. Then, the background of DTI is provided in Section 2.3.

2.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) scanning is medical imaging that uses a strong magnetic field and radio frequency waves to produce detailed images of the body structure. MRI scans can be used to diagnose many conditions (e.g. brain tumours, injuries, prostate problems and bone pathology). Figure 2.1 shows a closed MRI scanner. Unlike X-rays, an MRI scan does not use any ionizing radiation. To perform a study, the patient lies on a flat motorized bed that move inside the scanner either head or feet first depending on which part of the body needs to be scanned. A radiographer controls the scanner using a computer and communicates with the

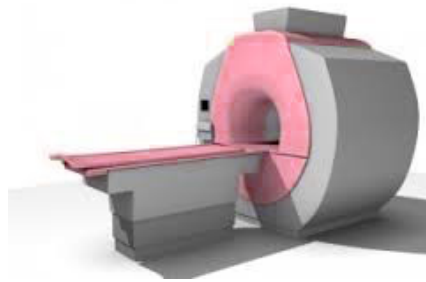


Figure 2.1: Closed MRI scanner

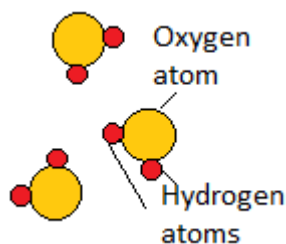


Figure 2.2: Water molecule consists of two hydrogen atoms and one oxygen atom.

patient through an intercom and a television monitor throughout the scan. Staying still during the scan is very important to avoid blurring the image.

How does an MRI scan work? Water is distributed almost everywhere in our bodies. In fact, our bodies are 50%–75% water. Each water molecule consists of two hydrogen atoms and one oxygen atom (Figure 2.2). The hydrogen atom with single proton (protium, ^1H), which is abundant in water and also in fat, is very sensitive to a magnetic field (Berger, 2002). Figure 2.3 shows ^1H which consists of an electron and proton. The proton spins with the magnetic field direction. Normally, in the absence of an external magnetic field, the protons (nucleus) are randomly oriented. When the body is in an MRI scanner, a strong magnetic field (B_0) will be applied into the body. The spin of the nuclei around B_0 is called precession. The rate of precession define the *precessional or Larmor frequency* which is directly proportional to the B_0 strength (Callaghan, 1991). There is a higher rate of precession for a stronger magnetic field. Slightly more than the half nucleus will be aligned parallel to B_0 direction (named as low energy nucleus), whereas the rest of the nucleus will be aligned anti-parallel (parallel with opposite direction) to B_0 direction (named as

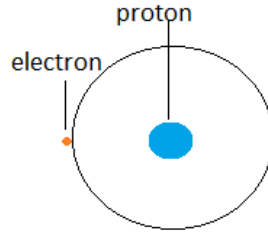


Figure 2.3: Hydrogen atom (protium, ^1H) consists of an electron and proton

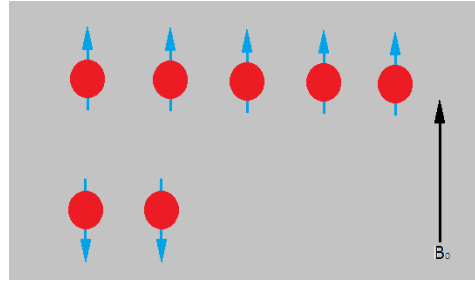


Figure 2.4: Proton spins parallel and anti-parallel to magnetic field direction.

high energy nucleus) (see Figure 2.4). This fact according to Boltzmann statistics that shows $N^-/N^+ = e^{-E/kT}$ (M. A. Brown & Semelka, 2011; Hornak, 2015) where N^- is the number of the low energy state protons, N^+ is the number of protons in the high energy state, E is the energy difference between the two states, k is Boltzmann's constant and T is the temperature.

Then, short bursts of radio frequency (RF) are sent from RF coil in the scanner into the body. The excited low energy nucleus will absorb the radio frequency energy and they align anti-parallel to the magnetic direction (i.e. changing their states from low energy to high energy). When the radio frequency stops, the nucleus that has recently changed their energy state, to high state, will emit the RF and return to low energy state. The interaction between the protons and the RF coil is called *resonance*. The *relaxation* process is when excited nuclei return to a low energy state. There are two independent relaxation processes: Spin-lattice and Spin-spin relaxation time (Callaghan, 1991; teaching.shu.ac.uk, 2015). Spin-lattice relaxation

($T1$ time) is related to losing the energy to the surrounding sample (lattice). Spin-spin relaxation ($T2$ time): interaction between the excited nuclei (in high energy state) and the neighbour nuclei in low energy leads to an exchange of the energy between them. In order to scan a specific region in the body, the region will be divided into slices. The RF will be sent with specific 'resonant' frequencies to each slice. The main magnetic field will be edited by *gradient coils*. There is a gradient coil in each of the three directions x, y and z to specify the slice orientation. The slice position and thickness will be determined depending on the gradient amplitude and RF (M. A. Brown & Semelka, 2011). One end of the coil has a stronger magnet and other end has a weaker magnet than the strength of B_0 . This variation is important to localize the scan into specific slices (howequipmentworks.com, 2012). After each slice is scanned, the receiver coil records the signal emitted by the nucleus and then the scan will continue to the next slice until all the region is scanned. The measured signal will be transformed to two different images: $T1$ and $T2$ weighted images. Tissues have different $T1$ and $T2$ times. For example fluids have longer $T1$ time (slow for returning to *relaxation*) so it appears dark in $T1$ image whereas fat has shorter $T1$ time and hence high signal (quicker for returning to *relaxation*) so it appears brighter (Morgan & Jones, 2015a). In $T2$ weighted image, fluids have bright colour and fat is intermediate bright coloured (Morgan & Jones, 2015b).

2.2 Diffusion Weighted Imaging

To define the diffusion weighted image we need to define first a voxel and Brownian motion. A *voxel* is an element in three-dimensional space picture. The voxel word is combination of “*volume*” and “*pixel*”. A *pixel* is an element in a two dimensional picture. Brownian motion is random movement of suspended particles in a fluid. It is named after Robert Brown 1827 who noticed this phenomena through the microscope (Ford, 1992). Albert Einstein (1905) published the article “On the Motion of Small Particles Suspended in a Stationary Liquid, as Required by the

Molecular Kinetic Theory of Heat” describing the Brownian motion model (Einstein, 1956). Einstein in 1956 explored the diffusion density f at time t in position x which follows the equation:

$$\frac{\partial f}{\partial t} = \mathbf{D} \frac{\partial^2 f}{\partial^2 x}, \quad (2.1)$$

where \mathbf{D} is the diffusion coefficient. The solution of Equation 2.1 provides the probability density function of molecular motion. Suppose there are n particles which start moving from the origin to position \mathbf{x} at initial time $t = 0$, then the probability density function of the n water molecules displacement is:

$$f(\mathbf{x}, t) = \frac{n}{\sqrt{4\pi\mathbf{D}t}} \exp\left(\frac{-\mathbf{x}^2}{4\mathbf{D}t}\right). \quad (2.2)$$

In DWI, the rate of water diffusion in each tissue voxel is measured. DWI uses the idea of Brownian motion to measure the mobility of the water molecule. Stejskal and Tanner (1965) introduce the magnetic field gradients by measuring the diffusion along specific directions called *diffusion gradient direction*. They reduced the MRI signal in regions with high diffusion. Le Bihan et al. (1986) introduce the parameter *b value* that works accurately in measuring the diffusion gradient along with magnetic gradient fields. Mansfield (1977) proposed the EPI technique that enables reducing imaging time. Mansfield shared the 2003 Nobel prize in medicine for his pioneering work towards MRI development.

2.3 Diffusion Tensor Imaging

In this section, a background of diffusion tensors is provided. This includes basic concepts, estimation methods and representation of diffusion tensor as an ellipsoid. In Addition, anisotropy measures and the brain views are explored. The processes and the applications of DTI are also discussed.

2.3.1 Basic Concepts

Diffusion Tensor Imaging DTI is an advanced MRI technique that measures the Brownian displacements of water molecules in each voxel in the brain and provides a unique information about biological tissues in the brain (P. Basser, Mattiello, & Le Bihan, 1994; P. Basser & Jones, 2002). The tri-variate Gaussian distribution of a three-dimensional random vector \mathbf{r} with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ can be obtained by:

$$f(\mathbf{r}) = \frac{1}{\sqrt{(2\pi)^3 |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{r} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu})\right). \quad (2.3)$$

The probability of a water molecule moved from origin to be in position $\mathbf{x} = (x, y, z)$ at time t is obtained as follows (D. Alexander, 2005; P. Basser, Mattiello, & Le Bihan, 1994; Einstein, 1956):

$$\begin{aligned} f(\mathbf{x}, t) &= \frac{1}{\sqrt{(4\pi t)^3 |\mathbf{D}|}} \exp\left(-\frac{1}{4t} \mathbf{x}^T \mathbf{D}^{-1} \mathbf{x}\right). \\ &= \frac{1}{\sqrt{(2\pi \times 2t)^3 |\mathbf{D}|}} \exp\left(-\frac{1}{2} \mathbf{x}^T (2t\mathbf{D})^{-1} \mathbf{x}\right), \end{aligned}$$

which is a tri-variate Gaussian distribution (Equation 2.3) with mean $\boldsymbol{\mu} = \mathbf{0}$ and covariance $\boldsymbol{\Sigma} = 2t\mathbf{D}$, where \mathbf{D} is the diffusion tensor. The diffusion of water molecules is a three dimensional process which can not be described by a scalar but needs a 3×3 , symmetric, positive semi-definite, real matrix (Le Bihan et al., 2001). That is,

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

where x , y and z represent the diffusion directions. The diffusion tensor can be estimated from the MR signal that can be acquired by setting up the scanner to n gradient directions. Let S_i be the MR signal in the i^{th} diffusion gradient direction.

Then $S_i \geq 0$ can be obtained as follows (P. Basser, Mattiello, & Le Bihan, 1994):

$$S_i = S_0 \exp(-b \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i), \quad \text{for } i \in \{1, \dots, n\}, \quad (2.4)$$

where \mathbf{D} is the diffusion tensor, $\mathbf{g}_i = \begin{bmatrix} g_{ix} & g_{iy} & g_{iz} \end{bmatrix}^T$ is the diffusion gradient in the i^{th} direction, b is a known scanner gradient parameter which characterizes the gradient attenuation (sec/mm^2) and S_0 is the Magnetic Resonance signal without the diffusion weighting (i.e. $b = 0$).

2.3.2 Estimation of Diffusion Tensor

Least square estimation is one of the methods that can be used to estimate the diffusion tensor. It is based on minimizing the sum of square residuals. The residual is the difference between the observed value (the actual value) and the fitted value (the value provided by the model). The sum of square residuals, $f(\boldsymbol{\beta})$, is obtained as a function of the parameter $\boldsymbol{\beta}$. The objective is finding $\boldsymbol{\beta}$ that minimizes $f(\boldsymbol{\beta})$ as follow:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} (f(\boldsymbol{\beta})). \quad (2.5)$$

Here, $\boldsymbol{\beta}$ can be found by equating the slope of $f(\boldsymbol{\beta})$ to zero. Note: argmin refers to $\boldsymbol{\beta}$ that minimizes $f(\boldsymbol{\beta})$. Koay, Carew, Alexander, Bassar, and Meyerand (2006) studied linear and non-linear least square estimation of diffusion tensors. They also explored constrained linear and non-linear least square methods by applying the Cholesky decomposition to preserve the positive definiteness condition on the estimated tensors.

Linear least square estimation (LLS)

In LLS, $f(\boldsymbol{\beta})$ should be written as linear function of $\boldsymbol{\beta}$. Rewrite the model in Equation 2.4 when $S_0 \neq 0$ as:

$$\ln(S_i/S_0) = -b \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i, \quad \text{for } i = 1, \dots, n. \quad (2.6)$$

As in Koay et al. (2006), Equation 2.6 can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}, \quad (2.7)$$

where

$$\mathbf{y} = [\ln(S_1/S_0), \dots, \ln(S_n/S_0)]^T, \quad (2.8)$$

$$\mathbf{X} = -b \begin{bmatrix} g_{1x}^2 & g_{1y}^2 & g_{1z}^2 & 2g_{1x}g_{1y} & 2g_{1y}g_{1z} & 2g_{1x}g_{1z} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{nx}^2 & g_{ny}^2 & g_{nz}^2 & 2g_{nx}g_{ny} & 2g_{ny}g_{nz} & 2g_{nx}g_{nz} \end{bmatrix}$$

and

$$\boldsymbol{\beta} = \begin{bmatrix} D_{xx} & D_{yy} & D_{zz} & D_{xy} & D_{yz} & D_{xz} \end{bmatrix}^T \quad (2.9)$$

is the vector representation of \mathbf{D} , which is the parameter we aim to estimate. The sum of square residuals for LLS estimation can be formulated as:

$$f_{LLS}(\boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \quad (2.10)$$

The estimator $\boldsymbol{\beta}$ that minimizes Equation 2.10 can be obtained by: $\hat{\boldsymbol{\beta}} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y}$.

Non-linear least square estimation (NLS)

For NLS estimation, Equation 2.4 can be written as follows (Koay et al., 2006):

$$S_i = S_0 \exp(\mathbf{X}\boldsymbol{\beta}), \quad i = 1, \dots, n \quad (2.11)$$

and the least square function of the parameter $\boldsymbol{\beta}$ is

$$f_{NLS}(\boldsymbol{\beta}) = \frac{1}{2} \|S_i - S_0 \exp(\mathbf{X}\boldsymbol{\beta})\|^2 = \frac{1}{2} \sum_{i=1}^n \left[S_i - S_0 \exp\left(\sum_{j=1}^6 X_{ij}\beta_j\right) \right]^2. \quad (2.12)$$

Equation 2.12 is a non-linear equation of the parameter $\boldsymbol{\beta}$ and hence there is no explicit solution for $\boldsymbol{\beta}$, whose computation needs a minimization algorithm (e.g.

GaussNewton algorithm).

Constrained linear (CLLS) and non-linear (CNLS) estimation

For both methods, LLS and NLS, Koay et al. (2006) suggested using Cholesky decomposition to preserve the positive definiteness condition on estimated diffusion tensors. We have $\mathbf{D} = \mathbf{L}^T \mathbf{L}$ where \mathbf{L} is upper triangular matrix as follows:

$$\begin{aligned} \mathbf{L} &= \begin{bmatrix} L_1 & L_4 & L_6 \\ 0 & L_2 & L_5 \\ 0 & 0 & L_3 \end{bmatrix} \text{ where } L_1, L_2 \text{ and } L_3 \neq 0. \text{ Then} \\ \mathbf{D} &= \begin{bmatrix} L_1 & 0 & 0 \\ L_4 & L_2 & 0 \\ L_6 & L_5 & L_3 \end{bmatrix} \begin{bmatrix} L_1 & L_4 & L_6 \\ 0 & L_2 & L_5 \\ 0 & 0 & L_3 \end{bmatrix} \\ &= \begin{bmatrix} L_1^2 & L_1 L_4 & L_1 L_6 \\ L_1 L_4 & L_2^2 + L_4^2 & L_4 L_6 + L_2 L_5 \\ L_1 L_6 & L_4 L_6 + L_2 L_5 & L_3^2 + L_5^2 + L_6^2 \end{bmatrix}. \end{aligned}$$

The vector representation of \mathbf{D} can be written as:

$$\boldsymbol{\beta} = \begin{bmatrix} L_1^2 & L_2^2 + L_4^2 & L_3^2 + L_5^2 + L_6^2 & L_1 L_4 & L_4 L_6 + L_2 L_5 & L_1 L_6 \end{bmatrix}^T \quad (2.13)$$

The sum of square residuals for CLLS and CNLS are:

$$f_{CLLS}(\boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2. \quad (2.14)$$

$$f_{CNLS}(\boldsymbol{\beta}) = \frac{1}{2} \|S_i - S_0 \exp(\mathbf{X}\boldsymbol{\beta})\|^2. \quad (2.15)$$

Minimizing equations 2.14 and 2.15 can be performed using any unconstrained, optimization algorithm (Koay et al., 2006).

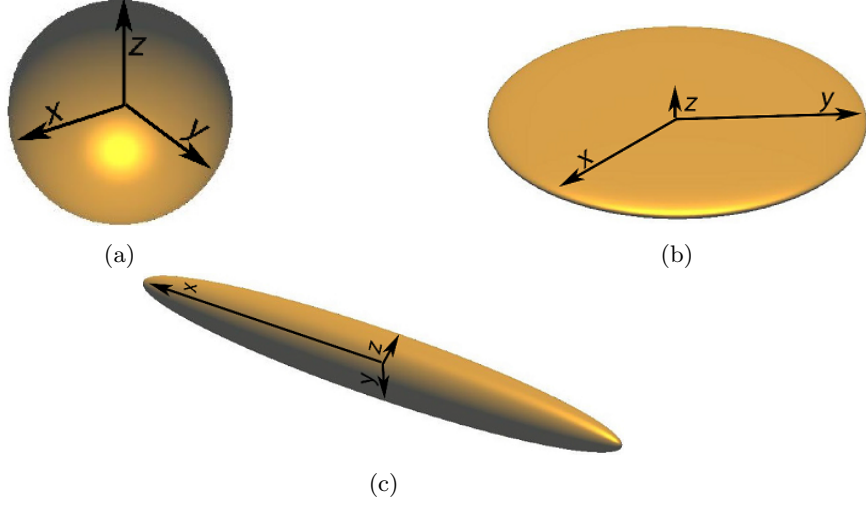


Figure 2.5: Isotropic diffusion (a) represented by spherical ellipsoid, whereas oblate (b) and cigar (c) shapes develop from anisotropic diffusion. The semi axis x , y and z represent the eigenvectors' direction, and the lengths of x , y and z are proportional to the square root of the eigenvalues of the tensor.

2.3.3 Eigensystem and Ellipsoid

A diffusion tensor can be expressed using its eigenvalues $\lambda_i \geq 0$ which represent the water diffusivities and eigenvectors \mathbf{v}_i , $i \in \{1, 2, 3\}$, which are parallel to the main diffusion directions. The spectral decomposition of \mathbf{D} is given by representing \mathbf{D} in terms of its eigenvalues and eigenvectors as follows:

$$\mathbf{D} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top.$$

where

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \quad (2.16)$$

and

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}. \quad (2.17)$$

Here, λ_i , for $i \in \{1, 2, 3\}$, are greater than 0 in the case of a positive definite tensor, or at least one of the λ_i equals 0 in the case of a tensor with deficient rank. If $\lambda_1 \geq \lambda_2 \geq \lambda_3$ then v_1 corresponding to λ_1 is the principal eigenvector

of \mathbf{D} . A diffusion tensor can be represented as an ellipsoid whose axes are the eigenvectors and radii are proportional to the square root of the eigenvalues of the tensor (P. Basser, Mattiello, & Le Bihan, 1994). The shape of the ellipsoid reflects how anisotropic (i.e. directionally dependent) the diffusion is. A spherical ellipsoid, as in Figure 2.5(a), means that the diffusion is isotropic or random in all directions. In this case all of the eigenvalues should be approximately the same value, whereas an extreme anisotropy diffusion could be represented by oblate or cigar ellipsoid. An oblate shape (see Figure 2.5(b)) shows that two eigenvalues are approximately equal and much larger than the third eigenvalue. A cigar shape (Figure 2.5(c)) results from having the principal eigenvalue much larger than the other two eigenvalues.

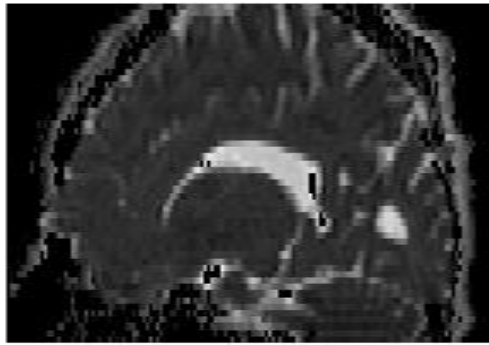
2.3.4 Diffusion Anisotropy Measures

To describe how much the water diffusion is anisotropic (or directional), several measures (indices) are proposed (Le Bihan et al., 2001). The most common used measures are Fractional Anisotropy, Relative Anisotropy and Volume Ratio. To define these measures, we need first to define the Mean Diffusivity (MD). MD measures the mean water diffusivity in each voxel in the brain and it does not provide any information about the diffusion direction. An example of MD map is shown in Figure 2.6a. The value of MD is equal to the mean of eigenvalues of the diffusion tensor:

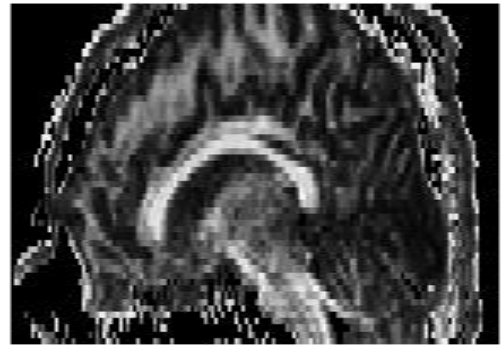
$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}.$$

Fractional Anisotropy (FA)

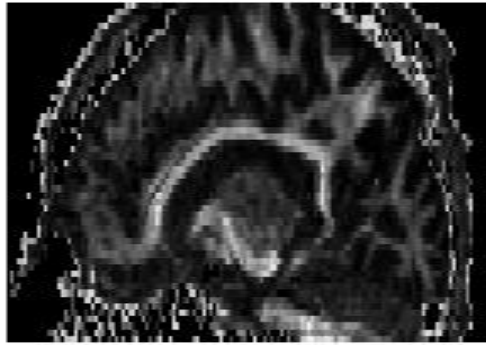
The FA measure (proposed by P. J. Basser and Pierpaoli (1996)) reflects the variability of water directions in each voxel in the brain. See Figure 2.6b for an example of FA map. Mainly, FA measures the deviation from isotropic diffusion using the standard deviation of the eigenvalues of \mathbf{D} . The FA values lie in $[0, 1]$, with $FA=0$ for isotropic diffusion and $FA=1$ for highly directional diffusion. In an FA map the dark regions provide low FA values whereas the bright regions provide high FA



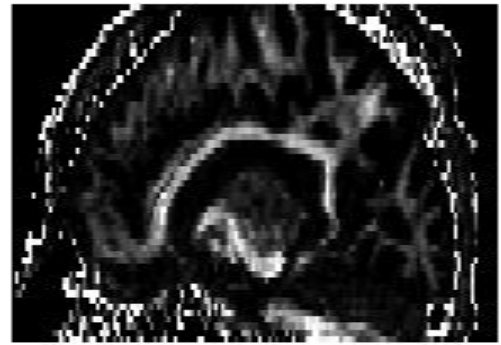
(a) Mean Diffusivity (MD) map



(b) Fractional Anisotropy (FA) map



(c) Relative Anisotropy (RA) map



(d) 1-Volume Ratio (VR) map

Figure 2.6: The sagittal brain view of Mean Diffusivity and anisotropy indices: FA, RA and 1-VR.

values. The FA value can be obtained as follows (P. J. Bassler & Pierpaoli, 1996):

$$FA = \sqrt{\frac{3[(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2]}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}},$$

which is equivalent to the formula:

$$FA = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}.$$

In the white matter of the brain, the diffusion is anisotropic due to the aligned fibre structures that restrict water diffusion in all directions (Mori, Wakana, Nagae-Poetscher, & Van Zijl, 2005). The diffusion tensor can be represented by an ellipsoid. A spherical ellipsoid means that the water diffuses equally in all directions and hence there is a high variability in water directions, which means low FA. An ellipsoid with a cigar shape means that the diffusion is directional and there is a less variability, which means high FA. For example, the FA for the spherical ellipsoid in Figure 2.5(a) is 0 whereas FA for the oblate ellipsoid (Figure 2.5(b)) and for the cigar ellipsoid (Figure 2.5(c)) are 0.6348 and 0.8911 respectively.

Relative Anisotropy (RA)

Another measure of anisotropy is the relative anisotropy (RA). An example of RA map is shown in Figure 2.6c. This measure is the ratio of the magnitude of anisotropic diffusion (measured by the standard deviation of eigenvalues) to magnitude of isotropic diffusion (measured by the mean of eigenvalues) (P. J. Bassler & Pierpaoli, 1996). The RA value can be obtained by:

$$RA = \frac{\sqrt{(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2}}{\sqrt{3}MD}.$$

Volume Ratio (VR)

The volume ratio (VR) can be defined as the ratio of the volume of the ellipsoid, whose semi axes are λ_1 , λ_2 and λ_3 , to the volume of the sphere whose radius is MD

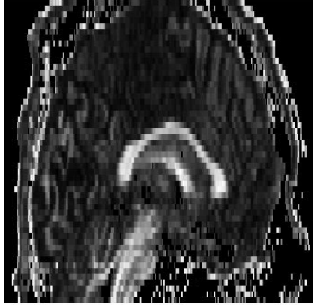
(Pierpaoli & Basser, 1996). The volume of the ellipsoid is $(4\pi/3)\lambda_1\lambda_2\lambda_3$ and the volume of the sphere is $(4\pi/3)MD^3$ and hence VR can be obtained by:

$$VR = \frac{\lambda_1\lambda_2\lambda_3}{MD^3}.$$

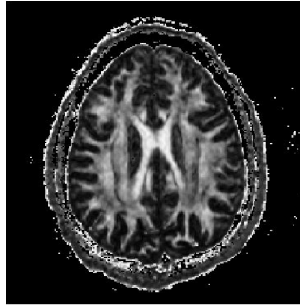
As the volume of ellipsoid increase, VR increases. We have $VR \in [0, 1]$, and is equal to 1 for spherical ellipsoid (isotropic diffusion) and 0 for high anisotropic diffusion. Then, $1 - VR$ is equivalent to FA (see Figure 2.6d for an example of $1 - VR$ map).

2.3.5 The Views of the Brain

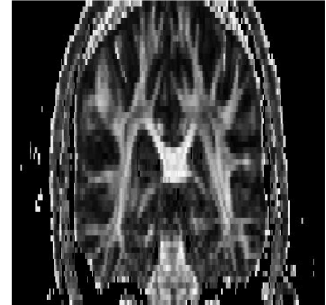
There are only three possible views of the brain: axial, sagittal and coronal. The axial view of the brain is an image of the brain taken along x and y axis (Figure 2.7a). The sagittal view of the brain is an image of the brain taken along y and z axis. If the sagittal image is taken from the middle of the brain, then the brain image is called midsagittal image (Figure 2.7b). The coronal view of the brain is an image of the brain taken along x and z axis (Figure 2.7c).



(a) The FA map of the midsagittal image of the brain



(b) The FA map of the axial view of the brain



(c) The FA map of the coronal view of the brain

Figure 2.7: The FA map of midsagittal, axial and coronal images of the brain

2.3.6 DTI Processing

There are many important processing methods to manipulate diffusion tensor data. For example: interpolation, smoothing and segmentation which will be explained with one example each, as follows.

Interpolation

Interpolation is estimating the values between two or more known values. Extrapolation involves estimating values outside (extending) the range of the known points. There are different types of interpolation: linear, polynomial, spline interpolation. Data resolution in DTI is sometimes weak, leading to missing values, which makes interpolation of diffusion tensors an essential process. Interpolation is also important for fiber tracking as there is a need for complete paths of tensors. One type of interpolation of DTI is geodesic interpolation which will be discussed next.

Geodesic interpolation

This method uses the weighted mean to interpolate a path between two diffusion tensors, named as a “geodesic path”. The weighted mean can be obtained using the Fréchet mean (Dryden et al., 2009). Suppose we need to interpolate a tensor Σ using n diffusion tensors \mathbf{D}_i , $i = 1, \dots, n$. Each \mathbf{D}_i has weight w_i . The interpolated tensor $\hat{\Sigma}$ can be obtained as follows:

$$\hat{\Sigma} = \arg \inf_{\Sigma} \sum_{i=1}^n w_i d^2(\mathbf{D}_i, \Sigma), \quad (2.18)$$

where $d(\mathbf{D}_i, \Sigma)$ is the distance between \mathbf{D}_i and Σ (could be an Euclidean or a non-Euclidean distance). The weight w_i should satisfy $w_i \geq 0$ and $\sum_{i=1}^n w_i = 1$. Generally there might be different choices of the weight function depending upon on the application or prior information. One choice of weight function could be using the inverse Euclidean distance as in (Dryden et al., 2009; Zhou, Dryden, Koloydenko,

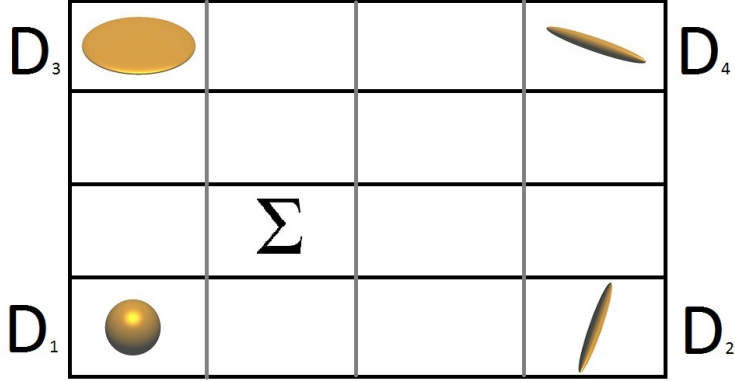


Figure 2.8: Interpolating the tensor Σ using four diffusion tensors \mathbf{D}_1 , \mathbf{D}_2 , \mathbf{D}_3 and \mathbf{D}_4 . Using geodesic interpolation, the tensor \mathbf{D}_1 will have the highest weight in determining $\hat{\Sigma}$ value as it is the closest to Σ .

& Bai, 2013). Zhou et al. (2013) suggested two different weight functions:

$$w_i = \frac{d_i}{\sum_{j=1}^n d_j}, \quad i \in \{1, \dots, n\}, \text{ and} \quad (2.19)$$

$$w_i = \frac{\exp(-Ad_i^2) + B}{\sum_{j=1}^n [\exp(-Ad_j^2) + B]}, \quad i \in \{1, \dots, n\}, \quad (2.20)$$

where d_i is the Euclidean distance from the position of the interpolated tensor to the position of the tensor \mathbf{D}_i , and $A, B > 0$ can be adapted depending on the application. That makes the exponential weight function in Equation 2.20 more flexible. For example, suppose we interpolate the tensor Σ in Figure 2.8 using the four diffusion tensors on the square corners $\mathbf{D}_1, \dots, \mathbf{D}_4$. We take $d_i, i \in \{1, \dots, 4\}$ to be the Euclidean distance from \mathbf{D}_i to Σ . As \mathbf{D}_1 is the closest to Σ , \mathbf{D}_1 will have the highest weight and hence most of $\hat{\Sigma}$ properties are derived from \mathbf{D}_1 . Then \mathbf{D}_2 and \mathbf{D}_3 have the same weight which is less than the weight of \mathbf{D}_1 . A smaller weight will be assigned to \mathbf{D}_4 and hence $\hat{\Sigma}$ will reflect fewer characteristics of \mathbf{D}_4 .

Smoothing Diffusion tensors

Smoothing the DTI is the process of removing the noise from the DTI. In fact, DTI provides low signal to noise ratio in some applications (e.g. high b value) which

leads to noisy images (D. Jones & Bassar, 2004). In those cases smoothing DTI is an essential process. Different methods of smoothing have been explored. Some of the previous methods smooth the diffusion weighted data whereas the others argue that smoothing the tensor field is the correct choice. Tabelow, Polzehl, Spokoiny, and Voss (2008) stated that smoothing the raw diffusion weighted image is superior to smoothing diffusion tensors. They justified that with two reasons. Firstly, smoothing the DWI provides less probability of estimating tensor with values outside the tensor space. Secondly, smoothing the DWI will reduce the bias in estimating the tensors which happens in case of high noise DWI. Smoothing using a weighted Fréchet (Karcher) mean (Equation 2.18), with any choice of Euclidean and non Euclidean means, is an example of smoothing a tensor field. One type of smoothing method is kernel smoothing, which can be used in either DWI smoothing or DTI smoothing. The kernel is a function which depends on the distance between two points and the bandwidth. The bandwidth is a parameter that controls the size of neighbourhood tensors that are used for smoothing. The kernel of the points \mathbf{x} and \mathbf{x}_i , with $i = 1, 2, \dots, n$, can be defined as:

$$f(\mathbf{x}, \mathbf{x}_i) = K_h(\|\mathbf{x}_i - \mathbf{x}\|), \quad \text{for } i = 1, \dots, n, \quad (2.21)$$

where

$$K_h(.) = (1/h)K(.), \quad h > 0. \quad (2.22)$$

The parameter h is called the kernel bandwidth and $K(.)$ is a non negative and integrable function that is either constant or inversely proportional to the Euclidean distance between \mathbf{x}_i and \mathbf{x} . An example of using kernel smoothing to smooth DTI follows.

Diffusion tensor smoothing through weighted Karcher means

This method has been explored by Carmichael, Chen, Paul, and Peng (2013). Like interpolation, the smoothing of diffusion tensors can be performed by using a weighted Fréchet (Karcher) mean. The weighted mean of the tensor neighbours can be used to estimate the new tensor. Carmichael et al. (2013) applied the smoothing using a weighted Karcher mean (Equation 2.18). They studied and compared the Euclidean, log Euclidean and Riemannian smoothing using a weighted Karcher mean. They used an anisotropic kernel to define the weight. Using an anisotropic kernel in smoothing DTI or DWI has been studied by many authors (e.g. Tabelow et al., 2008; Chung, Lee, & Alexander, 2005; J. Lee, Chung, Oakes, & Alexander, 2005). The following explanation has been paraphrased from Carmichael et al. (2013). The anisotropic kernel captures all the anisotropy information contained in the current diffusion tensor. Suppose the current estimate of the diffusion tensor is $\hat{\mathbf{D}}$ at position s and suppose $s_i, i \in \{1, \dots, n\}$ is the position of the tensor neighbours. The weight can be obtained by:

$$w_i(\mathbf{s}) := K_h(\sqrt{\text{tr}(\hat{\mathbf{D}})(\mathbf{s}_i - \mathbf{s})^T \hat{\mathbf{D}}^{-1}(\mathbf{s}_i - \mathbf{s})}), \quad \text{for } i = 1, 2, \dots, n,$$

where $K_h(\cdot)$ is defined in Equation 2.22. Using this method, Carmichael et al. (2013) performed two simulation studies and a brain image study. Both simulation studies and brain image suggested that all the methods (Euclidean, log Euclidean and Riemannian) smooth the isotropic regions, such as cerebrospinal fluid, similarly. Whereas, the log Euclidean and Riemannian methods effectively smooth high anisotropy regions such as the corpus callosum and fiber tracts in occipital lobe. The cerebrospinal fluid, Corpus Callosum (CC) and occipital lobe are shown in Figure 2.9. However, in the case where there is a variability in tensors structure due to DWI noise, Euclidean smoothing performs better than log Euclidean and Riemannian, as the latter produce an erroneous tensor structure. The method suggested the smoothing of diffusion tensors to be spatially adaptive to the geometric structures of the tensors and noise levels.

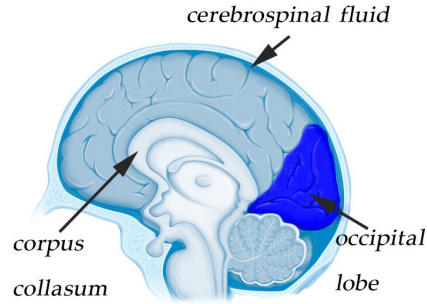


Figure 2.9: Corpus callosum, occipital lobe and cerebrospinal fluid

Segmentation of the CC

The CC is a great fiber bundle in the white matter of the brain. It connects the two hemispheres of the brain. Segmentation of the CC has very important clinical applications (e.g. surgical planning and disease diagnosis). Regarding disease diagnosis, the anatomical behavior of the CC in autism patients have been analyzed and it was shown that the CC size is smaller in patients with autism compared with healthy controls (Hardan et al., 2000; A. L. Alexander et al., 2007; El-Baz et al., 2011). In addition, the CC is smaller with high diffusivity for the temporal and frontal lobe of epilepsy patients (O'Dwyer et al., 2010). The microstructural anatomy of the CC is changed in early Huntington's disease (Rosas et al., 2010). Moreover, there is a significant reduction in FA in some subdivisions of the CC in cases of Multiple Sclerosis (Hasan, Gupta, Santos, Wolinsky, & Narayana, 2005), stroke (Gupta et al., 2006) and schizophrenia (Kubicki et al., 2008). Several works have been proposed towards the segmentation of the CC. Some research is concerned with segmenting the CC from the whole brain while others continue to segment the CC itself into subdivisions. Segmentation of the CC can be performed using MRI or DTI.

2.3.7 DTI Applications

Diffusion Tensor Imaging provides many important applications. It is important for following brain studies. For example for new born babies the anisotropy increases with age and hence decreasing anisotropy is an indicator of brain injury (S. P. Miller

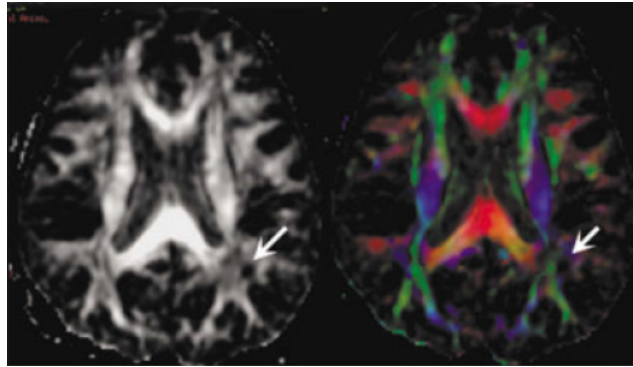


Figure 2.10: FA map (left) and color coded map (right) in a patient with chronic cerebral ischemia. There is a reduction in FA value in the lesion area (arrow). Taken from Dong et al. (2004). In the color coded map, blue represents superior-inferior fiber directions, green represents the anterior-posterior fiber directions and red represents the left-right fiber directions. Permission to reproduce this figure has been granted by John Wiley and Sons.

et al., 2002). In fact, DTI is more powerful in detecting some brain diseases at earlier stage than MRI (e.g. brain ischemia). Brain ischemia happens when the blood flow to the brain is insufficient which motivates the water flow from the extracellular compartment, that leads to swollen cells (Dong et al., 2004). The decrease in FA value in the lesion area is shown in Figure 2.10. In addition, DTI helps to detect brain ischemia in the early stages which is very important for effective therapy. Moreover, DTI is able to distinguish between acute and chronic ischemia which helps to improve the clinical care (Dong et al., 2004).

Regarding white matter diseases, DTI is able to detect and understand the pathophysiology of many white matter diseases such as multiple sclerosis. Werring, Clark, Barker, Thompson, and Miller (1999) demonstrate effectiveness of DTI in detecting multiple sclerosis showing that conventional MRI is not powerful to specify pathologic properties of the disease. The lesions in multiple sclerosis can be clearly seen in FA and MD map. In fact, DWI has already demonstrated its ability to detect many white matter diseases. For example, DWI helps to diagnosis the leukoencephalopathy syndrome (Ay et al., 1998). Separation of isotropy and anisotropy indices in DTI provides DTI with more powerful properties than DWI

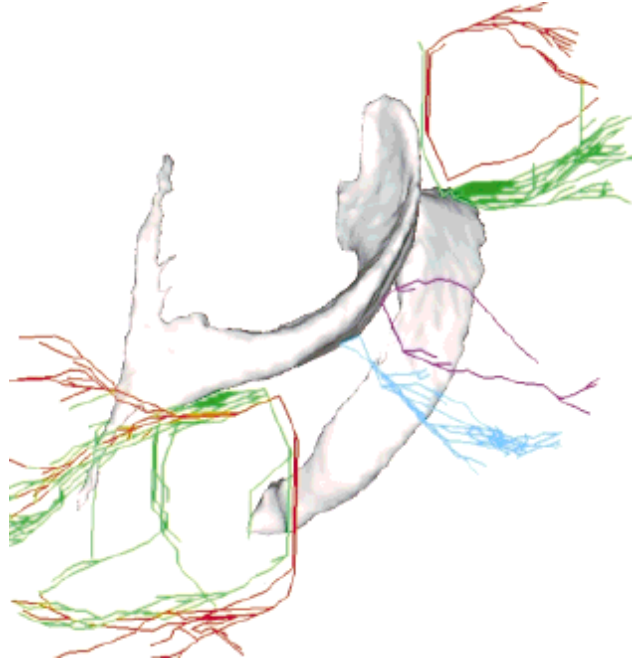


Figure 2.11: Fiber connectivity in the Callosum fibers that located around the brain ventricles. Taken from Le Bihan et al. (2001). Permission to reproduce this figure has been granted by John Wiley and Sons.

(Le Bihan et al., 2001).

How different parts of brain are connected can be described by brain connectivity (or fiber connectivity) which is one of the important applications of DTI. Brain connectivity helps to understand the anatomical structure of the brain. The main principal eigenvector represents the fiber orientation within a voxel unless fibers crossing or merging in a voxel (Dong et al., 2004). Figure 2.11 shows an example of fiber connectivity in the brain.

Chapter 3

Cluster Analysis of DTI

Segmentation of the Corpus Callosum (CC) plays an important role in disease diagnosis and surgical planning. We propose an automated method for segmentation of the CC using Hartigan's method and Diffusion Tensor Imaging. First, Hartigan's method is generalized to use with f -mean metrics (e.g. log Euclidean, root Euclidean and Cholesky) for clustering covariance matrices. This generalization is then applied to cluster diffusion tensor images of a healthy brain to segment the CC.

An introduction to this chapter is provided in Section 3.1. Then general background of the Euclidean and non-Euclidean methods as well as the processes using covariances matrices are presented in Section 3.2. Technical background of hierarchical clustering, K-means clustering and Hartigan's method is provided in Section 3.3. In Section 3.4, a generalized Hartigan's method that works for the f -mean metrics is proposed to provide an accurate clustering of diffusion tensors. Hierarchical clustering is used for a 2D segmentation of the CC and Hartigan's method is applied to cluster DTI to provide a 2D and 3D segmentation of the CC in Section 3.5. The hierarchical segmentation yields holes on the 2D CC and hence Hartigan's method is preferred. Using Hartigan's method, the log Euclidean and Riemannian methods provide the most accurate 2D segmentation when compared with manual segmentation. For 3D segmentation, the only methods that segments all of the splenium (posterior end) as part of the CC are the log Euclidean and Rie-

mannian methods. Hence, the log Euclidean and Riemannian methods outperform the other methods (i.e. Euclidean, Cholesky, root Euclidean and Procrustes) in the segmentation of the CC. We summarize and discuss the results in Section 3.6.

3.1 Introduction

The Corpus Callosum (CC) is the largest fiber bundle in the white matter of the brain. The CC has been demonstrated to be affected by many diseases: e.g. autism (Hardan et al., 2000; A. L. Alexander et al., 2007; El-Baz et al., 2011), multiple sclerosis (Hasan et al., 2005), stroke (Gupta et al., 2006), epilepsy (O’Dwyer et al., 2010), Huntington’s disease (Rosas et al., 2010) and schizophrenia (Kubicki et al., 2008). Specifically, the size and shape of the CC are affected by many brain disorders (e.g. see Duara et al., 1991; Hardan et al., 2000; Downhill et al., 2000; Walterfang et al., 2011; Bachman et al., 2014). Hence, segmentation of the CC is beneficial in the diagnosis of such diseases.

Magnetic Resonance Imaging (MRI) has been used widely for the segmentation of the CC (e.g. Lundervold, Duta, Taxt, & Jain, 1999; Jacob, Blu, & Unser, 2004; Sandhu, Georgiou, & Tannenbaum, 2008; El-Baz et al., 2011; Li, Mandal, & Ahmed, 2013). Diffusion Tensor Imaging (DTI) is an advanced MRI technique that measures the Brownian displacements of water molecules in each voxel in the brain and provides unique information about biological tissues in the brain (P. Basser, Mattiello, & Le Bihan, 1994; P. Basser & Jones, 2002). Segmentation of the CC has also been performed using DTI (e.g. S.-P. Lee, Cheng, Chen, & Tseng, 2008; Lenglet et al., 2006; Goh & Vidal, 2008b; Nazem-Zadeh et al., 2012). In S.-P. Lee et al. (2008), an automatic 2D segmentation approach of the CC using color coded map of DTI was proposed. The method proposed by Lenglet et al. (2006) was based on surface evolution which seeks the optimal partition through a Bayesian formulation. The method proposed by Goh and Vidal (2008b) used dimensionality reduction of diffusion tensors first and then performed traditional K -means clus-

tering. In Nazem-Zadeh et al. (2012), a three dimensional segmentation of the CC using DTI was proposed. They used the diffusivity pattern of the CC as prior information. A similarity measure, based on a speed function, has been proposed to segment the CC and its subdivisions.

A clustering of a finite data set is a partition of the data set into disjoint, nonempty subsets called clusters. Most clustering algorithms group the observations according to a similarity measure (i.e distance metric). Generally, there are three types of clustering: hierarchical clustering, partitioning and overlapping clusters (Seber, 2009). The K -means algorithm is an example of partitioning clustering and fuzzy c-means is an example of overlapping clustering. The only difference is that in Fuzzy c-means any object can belongs to more than one cluster. For every cluster, any object has a membership degree in the range $[0, 1]$ where 0 means this object is definitely not a member of this cluster, 1 means that the object is definitely a member of this cluster only and a value in between 0 and 1 means the object belongs to this cluster and at least one other cluster (Tan, Steinbach, Kumar, et al., 2006). That's because for any object the sum of the membership degree over all clusters is 1. There are a wide range of clustering applications (e.g. image processing, machine learning, bioinformatics and information retrieval). A K -means algorithm (Lloyd, 1982) is an unsupervised learning algorithm which aims to partition a data set into K clusters. Although, there are many K -means algorithms, all of them aim to minimize the within cluster sum of squares (WCSS). Hartigan's method (Hartigan, 1975) is a K -means algorithm which always provides smaller or equal WCSS as compared with the traditional Lloyd algorithm (Lloyd, 1982), and hence it is more accurate (Telgarsky & Vattani, 2010).

Non-Euclidean metrics (log Euclidean, Riemannian, Cholesky, Procrustes size-and-shape and root Euclidean) are alternatives to the Euclidean metric for the space of covariance matrices (Dryden et al., 2009). Non-Euclidean metrics are used for tensor processing such as regularization, interpolation and smoothing of tensor

fields (Pennec et al., 2006; Arsigny, Fillard, Pennec, & Ayache, 2007; Carmichael et al., 2013; Zhou, Dryden, Koloydenko, Audenaert, & Bai, 2016).

3.2 Background

3.2.1 Euclidean and Non-Euclidean Methods

A covariance matrix \mathbf{A} is a square symmetric matrix whose element in the i, j position is the covariance between i^{th} and j^{th} elements of a random vector \mathbf{v} . Let $\mathbf{v} = (x_1, \dots, x_n)$ and $E(x_i)$ is Expectation of x_i for $i \in \{1 \dots n\}$. As in A.V. Prokhorov (2011), the ij^{th} element in \mathbf{A}_n (i.e. $cov(x_i, x_j)$) is obtained as follows:

$$cov(x_i, x_j) = E[(x_i - E(x_i))(x_j - E(x_j))].$$

The elements in the main diagonal of a covariance matrix are variances as

$$cov(x_i, x_i) = E[(x_i - E(x_i))^2] = var(x_i).$$

Every covariance matrix is positive semi-definite (i.e the quadratic form $\mathbf{u}^T \mathbf{A} \mathbf{u} \geq 0$ for all $\mathbf{u} \in \mathbb{R}^n$). The space of covariance matrices is not an Euclidean space and so non-Euclidean measures might be more suitable for calculating the distance between two covariance matrices or the mean of a set of covariance matrices. There are different types of non-Euclidean distances and means which have been summarized in Dryden et al. (2009). The main structure follows Dryden et al. (2009) and paraphrased also from Arsigny et al. (2007), Pennec (2006) and Pennec et al. (2006). A brief summary of some Euclidean and non-Euclidean methods (i.e. distances and means) is provided in the following subsections.

Euclidean methods

Suppose that $\|\mathbf{A}\| = \sqrt{trace(\mathbf{A}^T \mathbf{A})}$. Then the Euclidean distance between covariance matrices \mathbf{A}_1 and \mathbf{A}_2 can be obtained by:

$$d_E(\mathbf{A}_1, \mathbf{A}_2) = \|\mathbf{A}_1 - \mathbf{A}_2\|.$$

The Euclidean mean estimator can be obtained by using both maximum likelihood and least square methods. The estimated mean is:

$$\hat{\Sigma}_E = \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i.$$

Arsigny et al. (2007) stated that the Euclidean mean is a *swelling* that is due to the polynomial interpolation of its determinant rather than monotonic interpolation.

Non-Euclidean methods

To estimate the mean of covariance matrices, the Fréchet or the Karcher mean which are based on minimization of the variance is used (Dryden et al., 2009). The Fréchet mean (Fréchet, 1948) is based on minimizing the variance globally whereas the Karcher mean (Karcher, 1977) is based on minimizing the variance locally. The Fréchet mean is unique for spaces that are sufficiently peaked (i.e. they have non-positive sectional curvature) and hence Fréchet mean is unique for Euclidean and non-Euclidean spaces (see Dryden et al. (2009) for more details). The Fréchet mean Σ , that minimizes the squared distances, can be obtained by:

$$\Sigma = \arg \inf_{\Sigma} \left(\frac{1}{2} \int d(\mathbf{A}, \Sigma)^2 f(\mathbf{A}) d\mathbf{A} \right).$$

Note: $\arg \inf_{\Sigma} (Z(\Sigma))$ means Σ that gives the greatest lower bound of $Z(\Sigma)$. For sample $\mathbf{A}_1, \dots, \mathbf{A}_n$ of covariance matrices, the Fréchet mean can be estimated by $\hat{\Sigma}$ that satisfies:

$$\hat{\Sigma} = \arg \inf_{\Sigma} \sum_{i=1}^n d(\mathbf{A}_i, \Sigma)^2. \quad (3.1)$$

The non-Euclidean methods are: Riemannian, log Euclidean, Cholesky, root Euclidean and Procrustes. The description of these methods is as follows. In the following, log without base refers to \log_e .

Riemannian method

Pennec et al. (2006) generalizes the mean, expectation and the variance measurements on a vector space onto a Riemannian manifold. In a connected Riemannian manifold, the distance between two points can be measured as the length of the shortest curve (geodesic) joining these points. In a Riemannian manifold (Pennec et al., 2006), the log function can be used to represent subtraction, i.e. the vector $\vec{\mathbf{p}\mathbf{q}}$ at the point p is equal to $\log_{\mathbf{p}}(\mathbf{q})$, whereas subtraction in a vector space is $\vec{\mathbf{p}\mathbf{q}} = \mathbf{q} - \mathbf{p}$. The exponential map in a Riemannian manifold represents addition i.e. $\mathbf{q} = \exp_{\mathbf{p}}(\vec{\mathbf{p}\mathbf{q}})$ (which means $p^{\vec{\mathbf{p}\mathbf{q}}} = \mathbf{q}$), whereas addition in a vector space is $\mathbf{q} = \mathbf{p} + \vec{\mathbf{p}\mathbf{q}}$. The Riemannian distance between two positive-definite covariance matrices \mathbf{A}_1 and \mathbf{A}_2 can be obtained by

$$d_R(\mathbf{A}_1, \mathbf{A}_2) = \left\| \log(\mathbf{A}_1^{-1/2} \mathbf{A}_2 \mathbf{A}_1^{-1/2}) \right\|, \quad (3.2)$$

where $\mathbf{A}_1^{-1/2}$ and $\log(\cdot)$ can be found using Equation 3.8 and 3.10 respectively. The mean estimator using the Fréchet mean (Equation 3.1) and the Riemannian metric (Equation 3.2) can be obtained by:

$$\hat{\Sigma}_R = \arg \inf_{\Sigma} \sum_{i=1}^n \left(\left\| \log(\mathbf{A}_i^{-1/2} \Sigma \mathbf{A}_i^{-1/2}) \right\|^2 \right).$$

Log Euclidean method

Using the logarithm of a matrix to define the distance between two matrices and the mean of a group of positive definite symmetric matrices has been explored by Pennec et al. (2006) and Arsigny et al. (2007). Arsigny et al. (2007) explained why using the Euclidean mean is not suitable for a positive definite space (i.e. especially for DTI). That is because, using the Euclidean mean the diffusion is interpreted by dispersion. This leads to a large determinant of the mean for large diffusion which

is not correct for DTI. The log Euclidean distance between two covariance matrices \mathbf{A}_1 and \mathbf{A}_2 is given by:

$$d_L(\mathbf{A}_1, \mathbf{A}_2) = \|\log(\mathbf{A}_1) - \log(\mathbf{A}_2)\|. \quad (3.3)$$

The estimator of the log Euclidean mean can be obtained by minimizing the sum of square log Euclidean distances between the covariance matrices. Using the Fréchet mean (Equation 3.1) and the log-Euclidean distance (Equation 3.3), the estimated mean is:

$$\hat{\Sigma}_L = \exp\left(\frac{1}{n} \sum_{i=1}^n \log \mathbf{A}_i\right),$$

where the exponential and the logarithm, $\exp(\cdot)$ and $\log(\cdot)$, can be obtained by using Equation 3.9 and 3.10 respectively.

Cholesky decomposition method

Cholesky factorization theorem: *If \mathbf{A} is a symmetric positive definite matrix then, there exists a unique factorization $\mathbf{A} = \mathbf{L}\mathbf{L}^T$ where \mathbf{L} is a lower triangular matrix with positive diagonal elements.* That is:

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{p1} \\ 0 & a_{22} & \cdots & a_{p2} \\ 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{pp} \end{bmatrix},$$

where a_{jj} , for $j = 1 \dots p$ are positive numbers. This Cholesky factorization can be extended to positive semi-definite matrices but the factorization is no longer unique. One application of Cholesky decompositions is modeling longitudinal data, the same type of information on the same variables at long periods of time (Dryden et al., 2009). Cholesky decomposition can be used as a way of re-parametrization of the

covariance matrix (Wang, Vemuri, Chen, & Mareci, 2004). The Cholesky distance between two covariance matrices \mathbf{A}_1 and \mathbf{A}_2 is given by:

$$d_C(\mathbf{A}_1, \mathbf{A}_2) = \|\mathbf{L}_1 - \mathbf{L}_2\|, \quad (3.4)$$

where $\mathbf{A}_i = \mathbf{L}_i \mathbf{L}_i^\top$, $i = 1, 2$ and $\mathbf{L}_i = \text{chol}(\mathbf{A}_i)$ is the Cholesky decomposition of \mathbf{A}_i . The mean of the covariance matrices can be estimated using the sample Fréchet mean (Equation 3.1) and the Cholesky distance (Equation 3.4) as:

$$\hat{\Sigma}_C = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{L}_i \right) \left(\frac{1}{n} \sum_{i=1}^n \mathbf{L}_i \right)^\top.$$

The use of Cholesky decomposition guarantees the positive definiteness characteristic of the estimated mean. The Cholesky mean provides an easy and quick computation of the mean estimator.

Root Euclidean method

Another method of re-parametrization of a covariance matrix is using the square root matrix, which has been explored by Dryden et al. (2009). The root Euclidean distance between two covariance matrices \mathbf{A}_1 and \mathbf{A}_2 is given by:

$$d_H(\mathbf{A}_1, \mathbf{A}_2) = \left\| \mathbf{A}_1^{1/2} - \mathbf{A}_2^{1/2} \right\|. \quad (3.5)$$

The estimator of Fréchet mean (Equation 3.1) using the root Euclidean distance (Equation 3.5) mean can be obtained by:

$$\hat{\Sigma}_H = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{A}_i^{1/2} \right) \left(\frac{1}{n} \sum_{i=1}^n \mathbf{A}_i^{1/2} \right)^\top$$

Procrustes size-and-shape method

The Procrustes method, named after a Greek bandit who forced people to fit in his

bed by chopping off their legs or stretching them, is used for analyzing sets of shapes. The Procrustean distance can be defined as "the shortest distance between sets of landmarks across all possible superimpositions" (N. M. Brown, 1994). Possible superimpositions mean translation, scaling and rotation. There are two Procrustes distances: full and partial (size-and-shape) Procrustes. In full Procrustes, translation, scaling, and rotation are performed to match objects together, whereas no scaling is performed for partial Procrustes distance. Here we are interested in Procrustes size-and-shape distance between two covariance matrices \mathbf{A}_1 and \mathbf{A}_2 which can be computed as follows:

$$d_S(\mathbf{A}_1, \mathbf{A}_2) = \inf_{\mathbf{R} \in O(n)} \|\mathbf{L}_1 - \mathbf{L}_2 \mathbf{R}\|, \quad (3.6)$$

where $\inf_{\mathbf{R}} f(\mathbf{R})$ means the greatest lower bound of $f(\mathbf{R})$, with respect to \mathbf{R} . To match \mathbf{L}_1 to \mathbf{L}_2 the Procrustes solution is

$$\hat{\mathbf{R}} = \arg \inf_{\mathbf{R} \in O(n)} \|\mathbf{L}_1 - \mathbf{L}_2 \mathbf{R}\| = \mathbf{U} \mathbf{V}^\top,$$

where $\mathbf{U}, \mathbf{V} \in O(n)$, the orthogonal group in dimension n , and $\mathbf{L}_1^\top \mathbf{L}_2 = \mathbf{V} \mathbf{\Lambda} \mathbf{U}^\top$ is the singular decomposition with diagonal matrix $\mathbf{\Lambda}$ of positive singular entries (Dryden et al., 2009). The sample Fréchet mean (Equation 3.1) using Procrustes size-and-shape distance (Equation 3.6) is:

$$\hat{\Sigma}_P = \hat{\mathbf{L}}_p \hat{\mathbf{L}}_p^\top,$$

where

$$\hat{\mathbf{L}}_p = \arg \inf_{\Sigma} \sum_{i=1}^n \left(\inf_{\mathbf{R}_i \in O(n)} \|\mathbf{L}_i \mathbf{R}_i - \Sigma\|^2 \right).$$

method	invariant1 metric	invariant2 metric	deficient rank	analytic mean solution
Euclidean	yes	no	can be used	exists
Riemannian	yes	yes	can't be used	does not exists
Log Euclidean	yes	yes	can't be used	exists
Cholesky	no	no	can be used	exists
Root Euclidean	yes	no	can be used	exists
Procrustes	yes	no	can be used	does not exists

Table 3.1: Comparison of the Euclidean and non-Euclidean methods

Comparison of the methods

In Table 3.1, invariant1 metric refers to the distance between \mathbf{A}_1 and \mathbf{A}_2 and whether it is invariant (unchanged), when rotation and reflection transformations are applied simultaneously to \mathbf{A}_1 and \mathbf{A}_2 . Invariant2 metric refers to the distance between \mathbf{A}_1 and \mathbf{A}_2 and whether it is invariant, under simultaneous scaling of \mathbf{A}_1 and \mathbf{A}_2 . Moreover, Riemannian distance is affine-invariant, that means: $d_R(\mathbf{A}_1, \mathbf{A}_2) = d_R(\mathbf{B}\mathbf{A}_1\mathbf{B}^\top, \mathbf{B}\mathbf{A}_2\mathbf{B}^\top)$ where $\mathbf{A}_1, \mathbf{A}_2$ and \mathbf{B} are $n \times n$ matrices and \mathbf{B} is a full rank matrix. The log Euclidean and Riemannian means are not suitable for rank deficient covariance matrices as the logarithm of 0 is an undefined value. All the mean estimators except the Riemannian and Procrustes mean are easy and quick to compute and there is no need to use numerical methods in their calculation. There is no explicit solution for the Riemannian and Procrustes mean estimators. For the Riemannian mean, the Gauss-Newton gradient descent algorithm has been proposed by Pennec (1991) and Pennec (2006) to estimate the mean value of the covariance matrices. For the Procrustes mean, the General Procrustes Algorithm is needed (Gower (1975); Dryden and Mardia (1998), page 90). The Euclidean, log Euclidean, Cholesky and root Euclidean means are straightforward to compute and do not require numerical methods. Generally, the Procrustes metric provides similar result to the root Euclidean metric. The maximum disagreement is $\sqrt{2}$ which happens when tensors are deficient with rank 1 (Zhou et al., 2016). In fact, the Procrustes metric outperforms the root Euclidean in some cases of rank deficiency (Pigoli, Aston, Dryden, & Secchi, 2014) (e.g. extrapolating in an infinite space). See

Dryden et al. (2009) and Zhou et al. (2016) for more details about non-Euclidean metrics. Note that we refer briefly to Procrustes size-and-shape as Procrustes.

3.2.2 DTI Processing Using Non-Euclidean Methods

Euclidean and non Euclidean means can be used for interpolation of the diffusion tensors. The interpolated tensors should have monotonic determinants, FA and MD (Yang et al., 2012). Non monotonicity in the determinants provides tensors with different sizes and non monotonicity in FA provides tensors with different shapes. Arsigny et al. (2007) stated that linear interpolation of determinants using the Euclidean mean is non monotonic which leads to big-sized tensors in the interpolated path (see Figure 3.1). Both the log Euclidean and Riemannian means are more effective in tensor interpolation than the Euclidean mean as the geometric interpolation of determinants is monotonic (Figure 3.1). The Cholesky mean (Wang et al., 2004) can be used for interpolation of the diffusion tensor, but it is unreliable in cases where the variability is small in particular for the diagonal elements of diffusion tensor (Dryden et al., 2009). Yang et al. (2012) argued that the orientation of the interpolated tensors using Euclidean, Cholesky and log Euclidean means are not considered and hence tensors with decreased FA and MD values are produced. Decreases in FA values might misguide the interpretation as decreases in FA values in white matter is an indicator of abnormalities (white matter disease) or crossing fibers (Yang et al., 2012). Two methods of interpolation based on Euler angles and the quaternion were developed in Yang et al. (2012) and their results with Euclidean, Cholesky, log Euclidean and Riemannian metrics were compared. The method demonstrated that the interpolated tensors are positive definite and provide monotonic determinant, FA and MD. The log Euclidean and Riemannian means are reported to be superior in smoothing a tensor field in the regions with high Fractional anisotropy (FA) in the brain compared with the Euclidean mean (Carmichael et al., 2013).

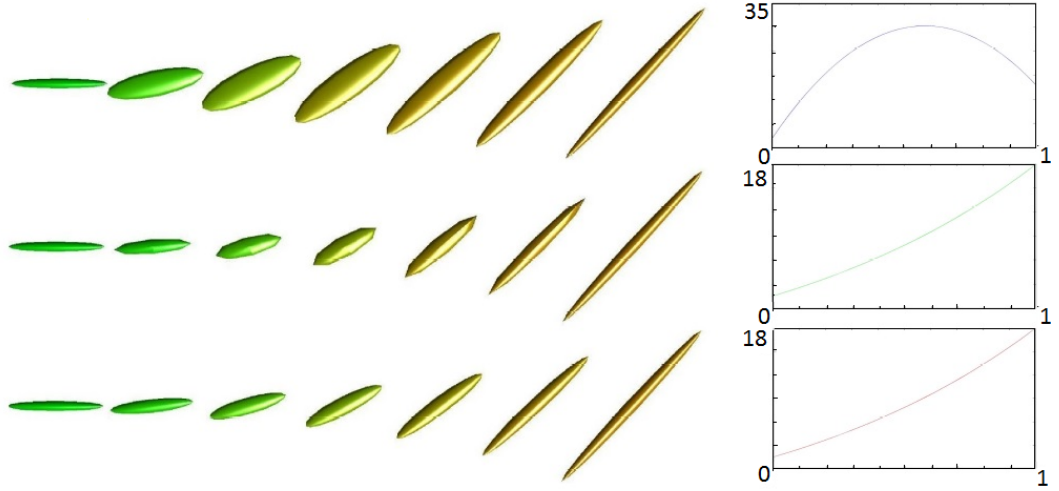


Figure 3.1: Geodesic interpolation path between two tensors. On the left side: Euclidean (top), Riemannian (middle) and log Euclidean (bottom) interpolation. On the right side: the corresponding determinants, which represent the monotonic interpolation of the determinants in the log Euclidean and Riemannian cases, and the non-monotonic interpolation of the determinants in the Euclidean mean (Arsigny et al. (2006)). Permission to reproduce this figure has been granted by John Wiley and Sons.

3.2.3 Segmentation of the CC Using Non-Euclidean Methods

The Riemannian metric has been used by both Lenglet et al. (2006) for CC segmentation. The Riemannian metric has been found to be more accurate for segmentation of the CC, especially on the splenium region, as compared with the Euclidean metric (Lenglet et al., 2006). Goh and Vidal (2008b) suggested an algorithm for segmentation of fiber bundles called a locally linear algorithm for diffusion tensor clustering (LLDTC). The algorithm first reduces the dimensionality of diffusion tensors to points, then clusters the points using a K -means algorithm and Riemannian metric. Recently Jayasumana et al. (2013) explored a kernel K -means for positive definite covariance matrices using Riemannian, log Euclidean, Cholesky and power Euclidean distances. Then they applied their method for Pedestrian Detection, Visual Object Categorization, Texture Recognition and segmentation of the CC. In fact the segmentation that has actually been provided in Jayasumana et al. (2013) is not a segmentation of the CC but a segmentation of the Lateral Ventricular (a

brain gray matter that is located under the CC).

The proposed segmentation method in this chapter might seem similar to the algorithm proposed by Goh and Vidal (2008b) in the way they both use DTI, K -means algorithm and Riemannian distance to segment the CC but they are completely different. The Goh and Vidal (2008b) method uses dimensionality reduction of diffusion tensors first and then traditional K -means clustering. In contrast, the proposed method clusters the diffusion tensors themselves without dimensionality reduction and hence all the information in the original data is used.

3.2.4 Processes on the Covariance Matrices

Consider some of the processes on the covariance matrix \mathbf{A} . The spectral decomposition of \mathbf{A} is given by $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$ where \mathbf{V} is an orthogonal matrix, $\mathbf{V} \in O(p)$, and $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues as follows:

$$\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_p \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_p \end{bmatrix}^\top, \quad (3.7)$$

where λ_i , for $i \in \{1, \dots, p\}$, are greater than 0 in the case of a positive definite covariance matrix, or at least one of the λ_i equals 0 in the case of a covariance matrix with deficient rank. Then \mathbf{A} to the power $k \in \mathbb{R}$ is obtained by:

$$\mathbf{A}^k = \mathbf{V}\mathbf{\Lambda}^k\mathbf{V}^\top. \quad (3.8)$$

The exponential of \mathbf{A} can be obtained by:

$$\exp(\mathbf{A}) = \sum_{t=0}^{\infty} \frac{\mathbf{A}^t}{t!} = \sum_{t=0}^{\infty} \frac{\mathbf{V}\mathbf{\Lambda}^t\mathbf{V}^\top}{t!} = \mathbf{V} \left(\sum_{t=0}^{\infty} \frac{\mathbf{\Lambda}^t}{t!} \right) \mathbf{V}^\top = \mathbf{V} \exp(\mathbf{\Lambda}) \mathbf{V}^\top. \quad (3.9)$$

To compute the logarithm of \mathbf{A} , \mathbf{A} should be a positive-definite covariance matrix (logarithm of 0 is an undefined value). The logarithm of \mathbf{A} is given by:

$$\log(\mathbf{A}) = \mathbf{V} \log(\mathbf{\Lambda}) \mathbf{V}^T, \quad (3.10)$$

where the entries of $\mathbf{\Lambda}$ are strictly positive eigenvalues ($\lambda_i > 0, \forall i$).

3.3 Cluster Analysis

In this section, a brief background to hierarchical clustering is provided, followed by description of K -means clustering and Hartigan's method.

3.3.1 Hierarchical Clustering

Hierarchical clustering is desirable in many applications for the structure that it provides e.g. in biological systems, medicine, psychology and network systems (Johnson, 1967; Ravasz, Somera, Mongru, Oltvai, & Barabási, 2002; Bandyopadhyay & Coyle, 2003). Unlike the K -means, hierarchical clustering does not depend on the cluster size. Moreover, hierarchical clustering is flexible as any distance metric can be used. There are two types of hierarchical clustering which can be explained as follow (Hastie et al., 2009):

- Agglomerative (bottom-up): Each object starts in its own cluster. As one moves up the hierarchy, the closest pair of clusters are merged at each level.
- Divisive (top-down): All objects start in one cluster. As one moves down the hierarchy, splits are performed recursively.

Given a set of n objects, a_i for $i \in \{1, \dots, n\}$, to be clustered, the basic process of agglomerative hierarchical clustering is shown below:

1. First assign each object to its own cluster, so there are n clusters, each containing one object.

2. Calculate the distances between all the clusters and obtain the square form i.e. $n \times n$ distance matrix Q as follows:

$$Q = \begin{matrix} & \begin{matrix} a_1 & a_2 & \cdots & a_n \end{matrix} \\ \begin{matrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{matrix} & \begin{pmatrix} 0 & d(a_1, a_2) & \cdots & d(a_1, a_n) \\ d(a_1, a_2) & 0 & \cdots & d(a_2, a_n) \\ \vdots & \vdots & \ddots & \vdots \\ d(a_1, a_n) & d(a_2, a_n) & \cdots & 0 \end{pmatrix} \end{matrix}, \quad (3.11)$$

where $d(a_i, a_j)$ is the distance between a_i and a_j , for $i, j \in \{1, \dots, n\}$. Look for the closest pair of objects and merge them into a single cluster, so there is now one less cluster ($n - 1$ clusters).

3. Compute distances between the new cluster and each of the old clusters. There are several ways to do this (e.g. single-link, complete-link and average-link). To calculate the distance between one cluster S_1 and another cluster S_2 , the methods (Seber, 2009) are:

- **Single-link clustering** (nearest neighbour): The distance between one cluster S_1 and another cluster S_2 is the **shortest** distance from any member of S_1 to any member of S_2 :

$$d_{S_1 S_2} = \min \{d_{rl}\},$$

where $r \in S_1$ and $l \in S_2$.

- **Complete-link clustering** (furthest neighbour): The distance between S_1 and S_2 is the **maximum** distance from any member of S_1 to any member of S_2 :

$$d_{S_1 S_2} = \max \{d_{rl}\}$$

- **Average-link clustering**: The distance between S_1 and S_2 is the **average** distance from any member of one cluster to any member of the

other cluster.

$$d_{S_1 S_2} = P(\bar{x}_{S_1}, \bar{x}_{S_2}),$$

where P is a proximity measure (e.g. square Euclidean, correlation), $\bar{x}_{S_1} = (1/n) \sum_{i \in S_1} x_i$ is the centroid of S_1 and similarly \bar{x}_{S_2} is the centroid of S_2 .

4. Repeat steps 2 and 3 until all objects are clustered into one cluster of size n .
5. Create Clusters:

After the hierarchical tree (also called dendrogram) has been created, it is required to determine where to cut the dendrogram into clusters. Using the **cluster** function in Matlab (Uk.mathworks.com, 2015a), there are two options to create the clusters:

- Inconsistency coefficient: the height of a link will be compared with the average height of the links below it, to obtain the inconsistency coefficient. By determining a specific value for tree inconsistency, the inconsistency coefficient can be used to divide the data into distinct, well-separated clusters.
- Specify the number of clusters that are required.

Suppose we need to cluster the objects: a, b, c, d and e using the single link method. In Figure 3.2, the agglomerative hierarchical clustering starts with five clusters and the closest pair of clusters will be merged at each level. At the final level there is one cluster for all the five objects. On the other hand, the divisive hierarchical clustering starts with one cluster for all the five objects and then clusters will be split recursively at each level. At the final level there are five clusters for the five objects. The height of a link in the dendrogram represents the distance between the objects. Suppose the distances between the five objects in Figure 3.2 as follow:

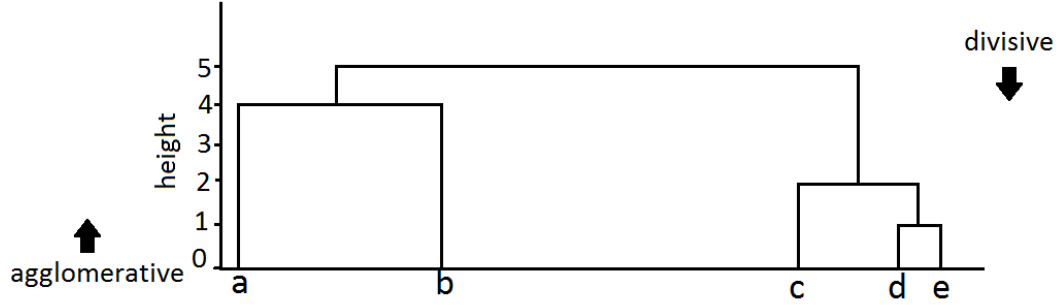


Figure 3.2: An example of the dendrogram for the hierarchical clustering (single link) using Equation 3.12.

$$Q = \begin{matrix} & a & b & c & d & e \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{pmatrix} 0 & 4 & 9 & 11 & 12 \\ 4 & 0 & 5 & 7 & 8 \\ 9 & 5 & 0 & 2 & 3 \\ 11 & 7 & 2 & 0 & \textcolor{red}{1} \\ 12 & 8 & 3 & \textcolor{red}{1} & 0 \end{pmatrix} \end{matrix} \quad (3.12)$$

The minimum distance is between d and e , and hence d and e will be grouped in one cluster. Then we need to calculate the distances between the new cluster (d, e) and other clusters as follows:

$$Q = \begin{matrix} & a & b & c & (d, e) \\ \begin{matrix} a \\ b \\ c \\ (d, e) \end{matrix} & \begin{pmatrix} 0 & 4 & 9 & 11 \\ 4 & 0 & 5 & 7 \\ 9 & 5 & 0 & \textcolor{red}{2} \\ 11 & 7 & \textcolor{red}{2} & 0 \end{pmatrix} \end{matrix} \quad (3.13)$$

For example, the distance between the cluster (d, e) and c using the single link method is $\min \{distance(d, c), distance(e, c)\}$. As the resultant minimum distance

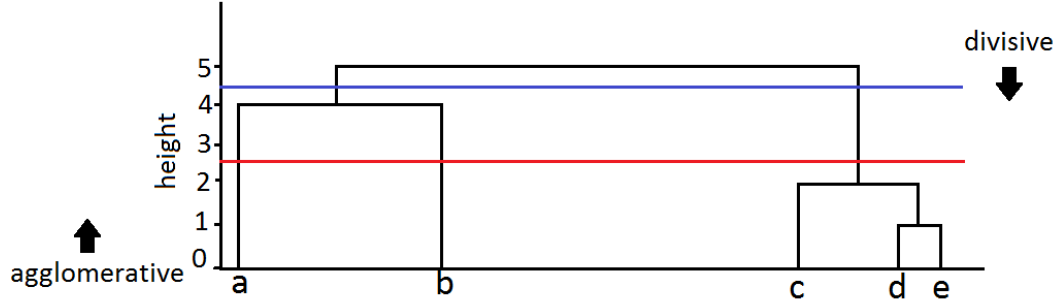


Figure 3.3: Cutting the dendrogram to obtain the clusters. The blue line cuts the dendrogram to obtain two clusters and the red line cuts the dendrogram to create three clusters.

is between c and (d, e) , they will be merged to obtain (c, d, e) :

$$Q = \begin{matrix} & \begin{matrix} a & b & (c, d, e) \end{matrix} \\ \begin{matrix} a \\ b \\ (c, d, e) \end{matrix} & \begin{pmatrix} 0 & 4 & 9 \\ 4 & 0 & 5 \\ 9 & 5 & 0 \end{pmatrix} \end{matrix} \quad (3.14)$$

Similarly, a and b will be merged into one cluster as follows:

$$Q = \begin{matrix} & \begin{matrix} (a, b) & (c, d, e) \end{matrix} \\ \begin{matrix} (a, b) \\ (c, d, e) \end{matrix} & \begin{pmatrix} 0 & 5 \\ 5 & 0 \end{pmatrix} \end{matrix} \quad (3.15)$$

Finally, the last two cluster will be merged together into one cluster which groups all the objects. To create the clusters we need to cut the dendrogram specific point (depending on an inconsistency coefficient or a given number of clusters). In Figure 3.3, the blue line cuts the dendrogram to obtain two clusters, (a, b) and (c, d, e) , whereas cutting the dendrogram using the red line creates three clusters a, b and (c, d, e) .

Hierarchical clustering has many advantages. One of the advantages is there is no need to specify the cluster size as input for the algorithm to work. In addition,

there is no need for the observations as the distances between the observations only are needed, and hence any distance metric can be used. However, the interpretation of the hierarchy is not easy to understand (Augen, 2004). Moreover, the Hierarchical clustering is less efficient (i.e. quadratic complexity) than the K -means clustering (linear complexity) (Tan et al., 2006).

Hierarchical clustering using non-Euclidean methods

The agglomerative hierarchical clustering (single link) can be used to cluster the covariance matrices using the non-Euclidean distances. Suppose we need to cluster n covariance matrices to different groups. We need the $n \times n$ distance matrix (Equation 3.11) using non-Euclidean distances (i.e. log Euclidean, Riemannian, Cholesky, root Euclidean and Procrustes). First assign each covariance matrix to its own cluster, so that there are n clusters each containing one covariance matrix. Look for the closest pair of covariance matrices and merge them into a single cluster, so that there are $n - 1$ clusters. In the single link method, the distances between the new cluster and the other clusters are computed as the shortest non-Euclidean distance from any covariance matrix in the new cluster to any member of the other cluster. We repeat calculating the distances and merging the closest pair until all the covariance matrices are a member of one cluster. Then we create clusters by cutting the dendrogram at specific point to obtain the required number of clusters.

3.3.2 Description of K -means

The most widely used and studied cluster analysis is the K -means clustering. In K -means clustering (MacQueen, 1967), we are given a set of n data points in d -dimensional space R^d , and the problem is to determine a set of $K \leq n$ points, centroids, so as to minimize the sum of the squared Euclidean distance from each data point to its nearest centroid (i.e. to minimize WCSS). The general procedure is to search for a K -partition, $C = \{C_1, \dots, C_K\}$, with locally optimal WCSS by moving points from one cluster to another. Given an integer k and a set \mathbf{x}_i , for $i \in \{1 \dots n\}$, the main objective is:

$$\arg \min_C \sum_{k=1}^K \sum_{\mathbf{x}_i \in C_k} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|^2, \quad (3.16)$$

where $\boldsymbol{\mu}_k$ is the Euclidean (arithmetic) mean of cluster C_k . The default distance metric is the Euclidean distance. Other distance metrics can be used in the K -means algorithm (e.g sum of absolute differences). The general procedure of K -means algorithm is described below.

K -means algorithm

Step 1: First select K random points as starting points acting as the centroids.

Step 2: Assign each point to its closest centroid.

Step 3: Recompute the centroid of each cluster.

Step 4: Repeat **Step 2** and **3** until all the points are nearest the centroid of the cluster to which they belong.

The K -means algorithm is simple and very efficient (linear complexity) but has some weaknesses. One of the weaknesses is that the K -means algorithm needs the number of clusters K to be specified. Moreover, the K -means algorithm minimizes WCSS and hence converges to a local minima. Thus, different initializations can lead to a different final clustering. This problem can be reduced by running the K -means algorithm with different initial partitions and selecting the partition with the smallest WCSS (Equation 3.16). In addition, K -means is not suitable for discovering clusters that has a non-globular (non-spherical) shapes (Tan et al., 2006). For example, Figure 3.4 shows two non-globular shapes where the K -means algorithm is not able to separate them into two different clusters (see Figure 3.5).

3.3.3 Hartigan's Method

Hartigan's method (Hartigan, 1975, p. 85-86) is a K -means algorithm to cluster a set of objects using Euclidean distance. In this section, we study Hartigan's method to cluster a set of matrices.

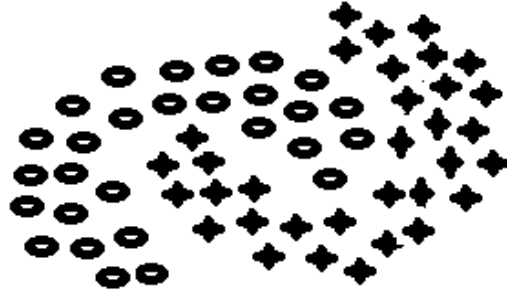


Figure 3.4: Two non-globular shapes

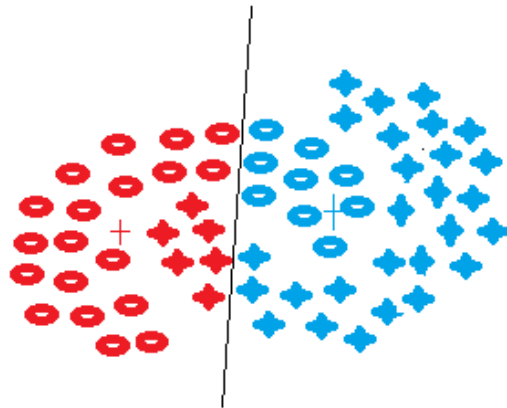


Figure 3.5: Two K -means clusters of the two non-globular shapes. The K -means algorithm fails to detect the two shapes and separate them into different clusters.

Basic notations

Suppose we need to cluster the set of matrices $\mathbf{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_M\}$ into K clusters. Suppose the initial partitioning $C = \{C_1, \dots, C_K\}$ where C_k , for $k \in \{1, \dots, K\}$, is the set of the indices of the matrices which clustered into cluster k . For example, suppose the set $A = \{A_1, \dots, A_6\}$ is clustered into $K = 2$ clusters with $\{A_1, A_4, A_5\}$ forming one cluster and $\{A_2, A_3, A_6\}$ forming the other cluster. Then $C = \{C_1, C_2\}$, with $C_1 = \{1, 4, 5\}$ and $C_2 = \{2, 3, 6\}$. Suppose $m(p)$, $p \in \{1, \dots, K\}$, is the number of matrices inside cluster p and $\bar{\mathbf{A}}_{C_p}$ is the centroid of cluster p . We refer to the within cluster sum of squares as $WCSS$, whereas $WCSS_p$ refers to the sum of squares within cluster p .

Description of Hartigan's method

Hartigan's method aims to minimize $WCSS$ by moving objects to another cluster if doing so will reduce $WCSS$. The Euclidean distance between \mathbf{A}_i and $\bar{\mathbf{A}}_{C_k}$ is defined as:

$$D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_k}) = \|\mathbf{A}_i - \bar{\mathbf{A}}_{C_k}\|,$$

where $\|\cdot\|$ is the Frobenius norm. Then $WCSS$ is defined as:

$$WCSS = \sum_{k=1}^K \sum_{i \in C_k} D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_k})^2. \quad (3.17)$$

Assume C_p and $C_j \in C$ and $i \in C_p$ which means the matrix \mathbf{A}_i is clustered into cluster p . Suppose \mathbf{A}_i is moved from cluster p into cluster j where $j \in \{1, \dots, K\}$ and $j \neq p$. Then $WCSS$ before the movement ($WCSS_{pre}$) is calculated as follows:

$$WCSS_{pre} = WCSS_1 + \dots + WCSS_p + \dots + WCSS_j + \dots + WCSS_K.$$

After the movement, $WCSS$ can be calculated as follows:

$$WCSS = WCSS_1 + \dots + WCSS_{np} + \dots + WCSS_{nj} + \dots + WCSS_K, \quad (3.18)$$

where $WCSS_{np}$ and $WCSS_{nj}$ are the new values of $WCSS_p$ and $WCSS_j$ after the movement. As in Spath (1985), $WCSS_{np}$ and $WCSS_{nj}$ can be calculated as

$$WCSS_{np} = WCSS_p - \frac{m(p)}{m(p) - 1} D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_p})^2 \quad (3.19)$$

and

$$WCSS_{nj} = WCSS_j + \frac{m(j)}{m(j) + 1} D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_j})^2. \quad (3.20)$$

Substituting Equation 3.19 and 3.20 into Equation 3.18 gives

$$\begin{aligned} WCSS &= WCSS_1 + \dots + WCSS_p - \frac{m(p)}{m(p) - 1} D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_p})^2 + \dots \\ &+ WCSS_j + \frac{m(j)}{m(j) + 1} D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_j})^2 + \dots + WCSS_K. \end{aligned}$$

Therefore

$$\begin{aligned} WCSS &= WCSS_{pre} - \frac{m(p)}{m(p) - 1} D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_p})^2 + \frac{m(j)}{m(j) + 1} D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_j})^2. \\ &= WCSS_{pre} + G_j, \end{aligned} \quad (3.21)$$

where

$$G_j = \frac{m(j)}{m(j) + 1} D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_j})^2 - \frac{m(p)}{m(p) - 1} D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_p})^2. \quad (3.22)$$

There are two possibilities:

- G_j is positive or zero, for every $j \neq p$. That is:

$$\forall j \in \{1, \dots, K\} \text{ such that } j \neq p, \text{ we have } G_j \geq 0. \quad (3.23)$$

This indicates that the movement of \mathbf{A}_i is unsuccessful (i.e. $G_j > 0$) or unnecessary (i.e. $G_j = 0$) as there is no reduction in $WCSS$, i.e. $WCSS \geq WCSS_{pre}$, or

- G_j is negative, for some j . That is:

$$\exists j \in \{1, \dots, K\} \text{ such that } G_j < 0. \quad (3.24)$$

That indicates the movement of \mathbf{A}_i is successful. In this case the movement will be to cluster j with the smallest G_j to achieve the largest reduction in $WCSS$, i.e. we want $\min_{j \neq p} G_j$.

3.4 Hartigan's method with f -mean

In this section we show how Hartigan's method can be applied using f -mean metrics to cluster a set of covariance matrices. First the f -mean is defined, followed by Hartigan's method using the f -mean.

3.4.1 F -mean

The f -mean is a generalization of the arithmetic, harmonic and geometric means. It is also known as the Kolmogorov (Kolmogorov & Castelnovo, 1930) or quasi-arithmetic mean. The f -mean is originally defined on the set of real numbers as a function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ (Tikhomirov, 2012). Here we define the f -mean to map between sets of covariance matrices. Suppose $f: \mathbb{I} \subset \mathbb{M} \rightarrow \mathbb{M}$, where \mathbb{M} denotes the class of covariance matrices, be a continuous and injective function. Then the f -mean for the covariance matrices $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$ is defined as:

$$\bar{\mathbf{A}} = f^{-1}\left(\frac{1}{n} \sum_{i=1}^n f(\mathbf{A}_i)\right). \quad (3.25)$$

The f -mean generalizes some common means such as the Euclidean, root Euclidean and power Euclidean mean, where $\mathbb{I} = \mathbb{M}$, and the log Euclidean and Cholesky mean where $\mathbb{I} = \mathbb{M}^+$ represents the class of positive-definite covariance matrices.

3.4.2 Hartigan's Method with the f -mean

Suppose we need to cluster a set of covariance matrices \mathbf{A} using Hartigan's method with f -mean metrics. Using the notation in Section 3.3.3, the $WCSS$ can be obtained by using Equation 3.17 where $D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_k})$ is defined as:

$$D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_k}) = \|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_k})\|. \quad (3.26)$$

From Equation 3.25, $f(\bar{\mathbf{A}}_{C_k})$ can be calculated as follows:

$$f(\bar{\mathbf{A}}_{C_k}) = \frac{1}{m(k)} \sum_{i \in C_k} f(\mathbf{A}_i),$$

and hence

$$\sum_{i \in C_k} f(\mathbf{A}_i) = m(k)f(\bar{\mathbf{A}}_{C_k}). \quad (3.27)$$

Suppose \mathbf{A}_i is moved from cluster p to cluster j . Let $\bar{\mathbf{A}}_{C_{np}}$ be the new centroid of cluster p after the movement. Let $C_{np} = C_p \setminus \{i\}$ and $C_{nj} = C_j \cup \{i\}$. The new centroid of cluster p can be calculated in terms of the old centroid as follows:

$$f(\bar{\mathbf{A}}_{C_{np}}) = \frac{1}{m(p) - 1} \sum_{l \in C_{np}} f(\mathbf{A}_l). \quad (3.28)$$

It is clear that

$$\begin{aligned} \sum_{l \in C_{np}} f(\mathbf{A}_l) &= \sum_{t \in C_p} f(\mathbf{A}_t) - f(\mathbf{A}_i) \\ &= m(p)f(\bar{\mathbf{A}}_{C_p}) - f(\mathbf{A}_i). \end{aligned} \quad (3.29)$$

Then substituting Equation 3.29 into Equation 3.28 yields:

$$f(\bar{\mathbf{A}}_{C_{np}}) = \frac{1}{m(p) - 1} [m(p)f(\bar{\mathbf{A}}_{C_p}) - f(\mathbf{A}_i)]. \quad (3.30)$$

Similarly, for cluster j ,

$$\begin{aligned} f(\bar{\mathbf{A}}_{C_{nj}}) &= \frac{1}{m(j)+1} \sum_{l \in C_{nj}} f(\mathbf{A}_l) \\ &= \frac{1}{m(j)+1} [m(j)f(\bar{\mathbf{A}}_{C_j}) + f(\mathbf{A}_i)]. \end{aligned} \quad (3.31)$$

The new centroids of clusters p and j can be obtained by taking the inverse function of $f(\bar{\mathbf{A}}_{C_{np}})$ and $f(\bar{\mathbf{A}}_{C_{nj}})$ in Equations 3.30 and 3.31 respectively. In Theorem 1, we show that Equations 3.19 and 3.20 and hence G_j in Equation 3.22 are applied for f -mean metrics (defined in Equation 3.26) as follows:

Theorem 1. *In Hartigan's method, suppose the covariance matrix \mathbf{A}_i is moved from cluster p to cluster j . Suppose the distance between two covariance matrices \mathbf{A}_1 and \mathbf{A}_2 is $\|f(\mathbf{A}_1) - f(\mathbf{A}_2)\|$, where $\|\cdot\|$ is the Frobenius norm and $\bar{\mathbf{A}}_{C_j} = f^{-1}(\frac{1}{m(j)} \sum_{t \in C_j} f(\mathbf{A}_t))$ is the f -mean of the covariance matrices in cluster j , then the function G_j for Hartigan's method is:*

$$G_j = \frac{m(j)}{m(j)+1} \|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})\|^2 - \frac{m(p)}{m(p)-1} \|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_p})\|^2. \quad (3.32)$$

Proof. For arbitrary $l \in C_{nj}$:

$$f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_{nj}}) = [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_j})] + [f(\bar{\mathbf{A}}_{C_j}) - f(\bar{\mathbf{A}}_{C_{nj}})]. \quad (3.33)$$

Using Equation 3.31 yields:

$$\begin{aligned} f(\bar{\mathbf{A}}_{C_j}) - f(\bar{\mathbf{A}}_{C_{nj}}) &= f(\bar{\mathbf{A}}_{C_j}) - \frac{1}{m(j)+1} [m(j)f(\bar{\mathbf{A}}_{C_j}) + f(\mathbf{A}_i)] \\ &= \frac{1}{m(j)+1} [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)]. \end{aligned} \quad (3.34)$$

Substitute Equation 3.34 into Equation 3.33 as follows:

$$f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_{nj}}) = [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_j})] + \frac{1}{m(j)+1} [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)]. \quad (3.35)$$

Equation 3.35 implies

$$\begin{aligned}
& [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_{nj}})] [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_{nj}})]^\top = \\
& \quad [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_j})] [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_j})]^\top \\
& \quad + \frac{2}{m(j)+1} [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_j})] [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)]^\top \\
& \quad + \frac{1}{(m(j)+1)^2} [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)] [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)]^\top.
\end{aligned} \tag{3.36}$$

To obtain $WCSS_{nj}$, we need to sum over $l \in C_{nj}$ for both sides of Equation 3.36, which is equivalent to taking the sum over $t \in C_j$ plus $l = i$ as follows:

$$\begin{aligned}
& \sum_{l \in C_{nj}} [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_{nj}})] [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_{nj}})]^\top = \\
& \quad \sum_{t \in C_j} [(f(\mathbf{A}_t) - f(\bar{\mathbf{A}}_{C_j}))(f(\mathbf{A}_t) - f(\bar{\mathbf{A}}_{C_j}))^\top] \\
& \quad + \underbrace{[f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})] [f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})]^\top}_{a_1} \\
& \quad + \underbrace{\frac{2}{m(j)+1} \sum_{t \in C_j} [(f(\mathbf{A}_t) - f(\bar{\mathbf{A}}_{C_j}))(f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i))^\top]}_b \\
& \quad + \underbrace{\frac{2}{m(j)+1} [f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})] [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)]^\top}_{a_2} \\
& \quad + \underbrace{\frac{m(j)+1}{(m(j)+1)^2} [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)] [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)]^\top}_{a_3}.
\end{aligned} \tag{3.37}$$

Adding terms together yields

$$\begin{aligned}
& a_1 + a_2 + a_3 = \\
& (1 - \frac{2}{m(j)+1} + \frac{1}{m(j)+1}) [f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})] [f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})]^\top \\
& = \frac{m(j)}{m(j)+1} [f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})] [f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})]^\top
\end{aligned} \tag{3.38}$$

and

$$\begin{aligned}
b &= \frac{2}{m(j)+1} [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)]^\top \left[\sum_{t \in C_j} [f(\mathbf{A}_t) - f(\bar{\mathbf{A}}_{C_j})] \right] \\
&= \frac{2}{m(j)+1} [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)]^\top \left[\left(\sum_{t \in C_j} [f(\mathbf{A}_t)] \right) - m(j)f(\bar{\mathbf{A}}_{C_j}) \right] \\
&= \frac{2}{m(j)+1} [f(\bar{\mathbf{A}}_{C_j}) - f(\mathbf{A}_i)]^\top [m(j)f(\bar{\mathbf{A}}_{C_j}) - m(j)f(\bar{\mathbf{A}}_{C_j})] = 0.
\end{aligned} \tag{3.39}$$

Then substitute Equation 3.38 and 3.39 in 3.37 as follows:

$$\begin{aligned}
&\sum_{l \in C_{nj}} [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_{nj}})] [f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_{nj}})]^\top \\
&= \sum_{t \in C_j} [(f(\mathbf{A}_t) - f(\bar{\mathbf{A}}_{C_j}))(f(\mathbf{A}_t) - f(\bar{\mathbf{A}}_{C_j}))^\top] + a_1 + a_2 + a_3 \\
&= \sum_{t \in C_j} [(f(\mathbf{A}_t) - f(\bar{\mathbf{A}}_{C_j}))(f(\mathbf{A}_t) - f(\bar{\mathbf{A}}_{C_j}))^\top] \\
&\quad + \frac{m(j)}{m(j)+1} [f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})] [f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})]^\top.
\end{aligned} \tag{3.40}$$

Then, we take the trace of Equation 3.40 as follows:

$$\begin{aligned}
\sum_{l \in C_{nj}} \|f(\mathbf{A}_l) - f(\bar{\mathbf{A}}_{C_{nj}})\|^2 &= \sum_{t \in C_j} \|f(\mathbf{A}_t) - f(\bar{\mathbf{A}}_{C_j})\|^2 \\
&\quad + \frac{m(j)}{m(j)+1} \|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})\|^2.
\end{aligned}$$

Therefore

$$WCSS_{nj} = WCSS_j + \frac{m(j)}{m(j)+1} \|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})\|^2. \tag{3.41}$$

In addition, the WCSS of cluster p , after the movement, is

$$WCSS_{np} = WCSS_p - \frac{m(p)}{m(p)-1} \|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_p})\|^2. \tag{3.42}$$

Then the change amount in WCSS, where $j \neq p$, is

$$G_j = \frac{m(j)}{m(j) + 1} \|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_j})\|^2 - \frac{m(p)}{m(p) - 1} \|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_{C_p})\|^2. \quad (3.43)$$

□

In our applications in Section 3.5, particularly for the 3D segmentation, the region of interest (ROI) is a very large volume of voxels from the middle of the brain, and hence efficiency is important. Hartigan's method runs very slowly for large data sets as it depends on moving objects one at a time. To improve efficiency, the centroids resulting from Lloyd's algorithm can be used as an initialization to Hartigan's method. For the methods that don't have an analytical solution for the computation of the mean, i.e. the mean cannot be computed using Equation 3.25, Hartigan's method is not applicable. However, the algorithm can be edited just to check if an existing partitioning result can be improved further (by finding a partitioning with a smaller *WCSS*) but not to start partitioning from the beginning. Hence the initialization can be performed using Lloyd's algorithm but not using a random partitioning. Furthermore, adjusting *WCSS* and the centroid values, after a movement, cannot be performed using the old values of *WCSS* and centroid before the movement. The generalized algorithm is shown below. For the f -mean, the following applies:

- The initialization in Line 1 could also be performed using random initialization, and the centroids of the clusters are calculated using Equation 3.25.
- Updating the centroids in Line 9 are performed by taking the inverse function in Equations 3.30 and 3.31.
- Equation 3.21 with the metric defined in Equation 3.26 is used to adjust *WCSS* in Line 10.
- Lines 12-17 are not needed.

Algorithm 1: Generalized Hartigan's method.

```
1 Initialization: use Lloyd's algorithm to partition  $\mathbf{A}$  into  $K$  groups as
   the initialization,  $C = \{C_1, \dots, C_K\}$ . Then compute the centroids of
   the clusters and  $WCSS$  (Equation 3.17).
2 while  $move = true$  do
3    $move = false$ ;
4   for  $i \in \{1, \dots, M\}$  do
5     Let  $i \in C_p$ , where  $p \in \{1, \dots, K\}$ .
6     if  $G_j < 0$  for some  $j \in \{1, \dots, K\}$  and  $j \neq p$  then
7       Let  $i \in C_j$ , where  $j = \arg \min_{j \neq p} G_j$ .
8       Let  $move = true$ .
9       Update the centroids of cluster  $p$  and  $j$ ,  $\bar{\mathbf{A}}_{C_{np}}$  and  $\bar{\mathbf{A}}_{C_{nj}}$ .
10      Let  $WCSS_{pre} = WCSS$ , then adjust the new  $WCSS$ .
11    end
12    if  $WCSS < WCSS_{pre}$  then
13      let  $move = true$ .
14    else
15      Undo the movement and restore the previous values of the
        centroids of cluster  $p$  and  $j$ .
16      Let  $WCSS = WCSS_{pre}$  .
17    end
18  end
19 end
20 Output:  $C = \{C_1, \dots, C_K\}$ .
```

For the methods that have a mean other than the f -mean, such as Procrustes and Riemannian, the following alterations of the algorithm are required:

- In Line 1 and 9, the centroids are calculated using the appropriate mean (e.g. see Appendix A.2 for the Procrustes and Riemannian means).

- Line 8 is not needed.
- In Line 10, the WCSS is adjusted using Equation 3.17 with the appropriate metric (e.g. see Appendix A.1 for the Procrustes and Riemannian distances).

3.4.3 Special Cases

Here we study Hartigan's method with some special cases of the f -mean i.e. log Euclidean, root Euclidean, Cholesky and Euclidean method. We show how the centroids can be adjusted, after a movement, in an easy and quick way using the old values before the movement.

When $f(\mathbf{A}) = \log(\mathbf{A})$, this is the case of the log Euclidean method. The log Euclidean distance between two covariance matrices \mathbf{A}_1 and \mathbf{A}_2 is

$$\|\log(\mathbf{A}_1) - \log(\mathbf{A}_2)\|.$$

The centroid of cluster p using the log Euclidean method can be obtained by substituting $f(\mathbf{A}) = \log(\mathbf{A})$ into Equation 3.25 as follows:

$$\bar{\mathbf{A}}_{C_p} = f^{-1}\left(\frac{1}{m(p)} \sum_{t \in C_p} \log(\mathbf{A}_t)\right) = \exp\left(\frac{1}{m(p)} \sum_{t \in C_p} \log(\mathbf{A}_t)\right).$$

Then

$$f(\bar{\mathbf{A}}_{C_p}) = \log(\bar{\mathbf{A}}_{C_p}) = \log\left(\exp\left(\frac{1}{m(p)} \sum_{t \in C_p} \log(\mathbf{A}_t)\right)\right) = \frac{1}{m(p)} \sum_{t \in C_p} \log(\mathbf{A}_t).$$

Hence, using Equations 3.30 and 3.31, the new centroids when \mathbf{A}_i is moved from cluster p to cluster j are respectively:

$$\bar{\mathbf{A}}_{C_{np}} = \exp\left(\frac{1}{m(p)-1} [m(p) \log(\bar{\mathbf{A}}_{C_p}) - \log(\mathbf{A}_i)]\right)$$

and

$$\bar{\mathbf{A}}_{C_{nj}} = \exp \left(\frac{1}{m(j) + 1} [n \log (\bar{\mathbf{A}}_{C_j}) + \log (\mathbf{A}_i)] \right).$$

In addition, $WCSS$ can be adjusted after the movement using Equation 3.21 where

$$G_j = \frac{m(j)}{m(j) + 1} \|\log (\mathbf{A}_i) - \log (\bar{\mathbf{A}}_{C_j})\|^2 - \frac{m(p)}{m(p) - 1} \|\log (\mathbf{A}_i) - \log (\bar{\mathbf{A}}_{C_p})\|^2.$$

Similarly we can adjust the centroids and calculate G_j for root Euclidean, Cholesky and Euclidean using $f(\mathbf{A}) = \mathbf{A}^{1/2}$, $f(\mathbf{A}) = \text{chol}(\mathbf{A})$ and $f(\mathbf{A}) = \mathbf{A}$ respectively (see Appendix A.3).

3.5 Applications

A set of MR images of a healthy brain is provided by the Academic Radiology Department of Queens Medical Center provided. The *MR* images were acquired using a *3T* Phillips Achieva scanner with a receive-only eight elements head coil. The diffusion tensor MR images were acquired using a spin echo, EPI, sequence with diffusion weighting gradients applied with a weighting factor of $b = 1000 \text{ s/mm}^2$. A number of 45 interleaved contiguous transaxial slices were acquired throughout the subject's head in a matrix of 112×112 (interpolated to 224×224) with an acquisition voxel size of $2 \times 2 \times 2.5 \text{ mm}$. The following specifications were used in the experiments but their knowledge is not required for this work. Using a parallel imaging SENSE factor of 2 and half *Fourier factor* of 0.75 resulted in an *echo time* (TE) of *56 ms* and *repetition time* (TR) of *9700 ms*. For each slice, the acquisition was repeated to acquire diffusion weighted images in 31 non-collinear directions, as well as one acquisition with no diffusion weighting ($b = 0$). The acquisition was repeated twice and the data averaged to improve the signal to noise ratio. To minimize the clustering time of the whole brain image, we select a ROI that includes the CC. Hartigan's method is used for the 2D and 3D segmentation of the CC. Lloyd's algorithm is used for the initialization to reduce the time of clustering.

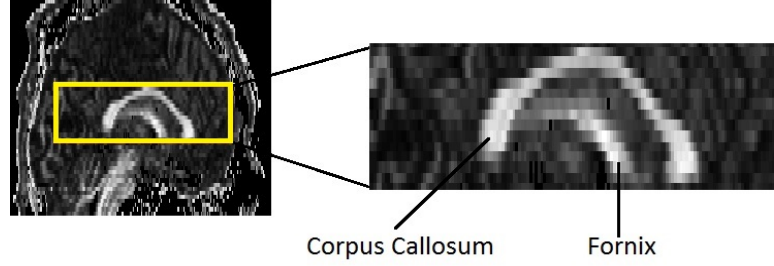


Figure 3.6: The Fractional Anisotropy (FA) map of the midsagittal plane of the brain. The region of interest (ROI) is highlighted by the yellow rectangle.

In the following subsections, we used hierarchical clustering and generalized Hartigan’s method to segment the CC of the healthy brain. Generally, holes are not expected in the healthy CC and hence we prioritized the criteria ‘no holes’, to compare between the segmentations using the hierarchical clustering and Hartigan’s method.

3.5.1 Segmentation using Hierarchical Clustering

Agglomerative hierarchical clustering (single-link) is used to segment the CC region in the brain using Euclidean and non-Euclidean methods. To reduce the clustering time of a whole slice, we choose a region of interest (ROI) from the midsagittal slice of the brain (Figure 3.6). Suppose the number of the tensors in ROI is n . First we calculate the square form ($n \times n$) of the non-Euclidean distances between the tensors. Then, we used the square form as input to Matlab *linkage* function (Uk.mathworks.com, 2015b), which has been used to create the hierarchical tree. Then the Matlab *cluster* function (Uk.mathworks.com, 2015a) is used to cluster the tree. The results of the first appearance of the CC shape are in Figure 3.7. When we increase the clusters size, we obtain a disjoint shape of the CC. Note that although the tensors in the fornix have high FA values like the tensors in the CC, hierarchical clustering differentiates well between them so the fornix is not included as part of the CC.

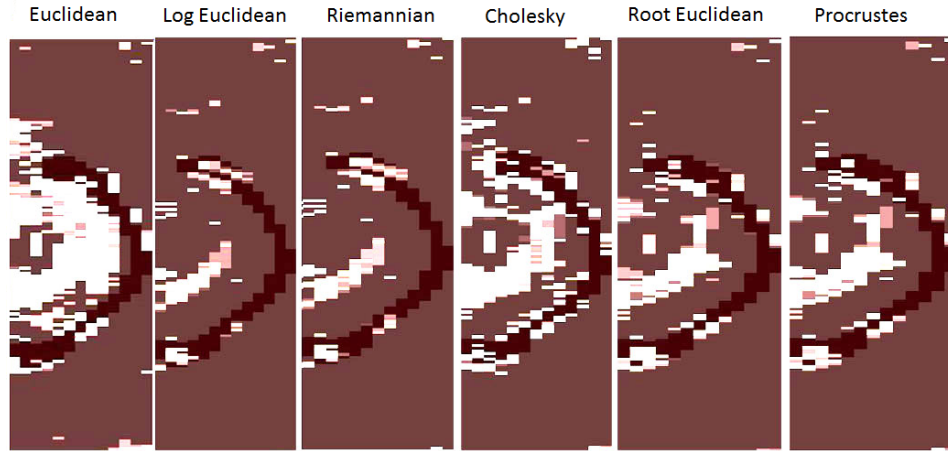


Figure 3.7: Hierarchical clustering using Euclidean and non-Euclidean methods to segment the CC in the brain using the midsagittal brain image.

Evaluation of the segmentation

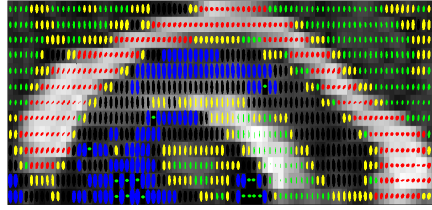
The CC shape is connected and holes on it is not expected in healthy subjects. Holes on the CC might indicate presence of lesions which is an indicator of abnormality (see Ho, Moonis, Ginat, and Eisenberg (2013)). The hierarchical (single-link) segmentation provides holes in the CC of healthy subjects. The method that provides the largest number of holes is not preferred as holes are not expected. To evaluate the segmentation, the number of holes inside the segmented shape of the CC (i.e. number of pixels that should be part of the CC shape but have been grouped into a different cluster), mean and standard deviation of the number of holes has been calculated in Table 3.2. The data used are five healthy subjects (the data from the healthy subject described in this chapter and the four healthy subjects described in Chapter 4). From Table 3.2, it can be seen that the Cholesky method yields the largest average number of holes in the CC whereas the Riemannian method produces the smallest average number of holes in the CC. Therefore the Riemannian method using the hierarchical clustering (single-link) outperforms the other methods for segmentation of the CC. Using the Euclidean method, the CC is segmented for three subjects and the CC is not visible for two subjects.

Method	No of holes					mean	std
	subject1	H1	H2	H3	H4		
Euclidean	5	16	NA	29	NA	16.67	12.01
Log Euclidean	25	4	23	15	30	19.40	10.16
Riemannian	25	2	14	15	27	16.60	10.01
Cholesky	18	9	17	19	37	20.00	10.30
Root Euclidean	18	13	27	14	26	19.60	6.58
Procrustes	18	12	17	23	28	19.60	6.11

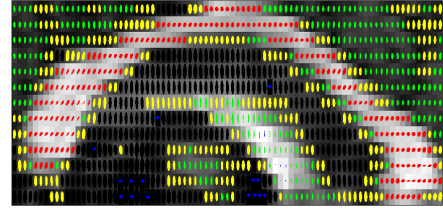
Table 3.2: The number of holes, mean number of holes and standard deviation of number of holes in the CC using hierarchical clustering (single link) for the healthy brain image described in this Chapter (subject1) and the four healthy subjects described in Chapter 4 (H1, H2, H3 and H4). Using the Euclidean method, the segmentation of the CC works only for three subjects (subject1, H1 and H3) and fails for two subjects (H2 and H4). The segmentation using the Riemannian method have the smallest average number of holes and hence it is preferred, while the segmentation using the Cholesky method has the largest average number of holes and hence it is not preferred.

3.5.2 Two Dimensional Segmentation Using Hartigan's Method

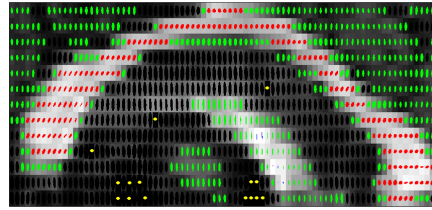
For $2D$ segmentation, the region of interest (ROI) from the midsagittal slice of the brain (shown in Figure 3.6) is used. Hartigan's method is used to cluster the tensors in the ROI into 5 clusters. How the cluster size is chosen is discussed below. The segmentation results with FA as the background are shown in Figure 3.8. The diffusion tensors for the 5 clusters are represented by diffusion ellipsoids in 5 colors: red (for the diffusion tensors in the CC), green, yellow, blue and black. Like the hierarchical clustering, Hartigan's method differentiates between the fornix and the CC. To confirm the clustering results, the variance of the CC using the same distance with different means is used. Using each of the Euclidean or non-Euclidean methods, the variance of the CC should be the smallest using the same distance and the same mean of the method (e.g. the smallest variance of the CC using the Euclidean distance and different means should be obtained using the Euclidean mean). Suppose that C_{cc} is the CC partition, let $m(cc)$ be the number of the tensors



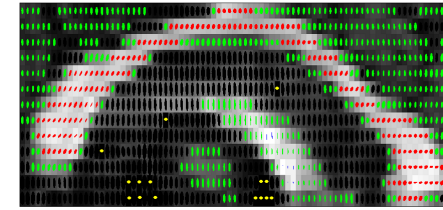
(a) Euclidean method



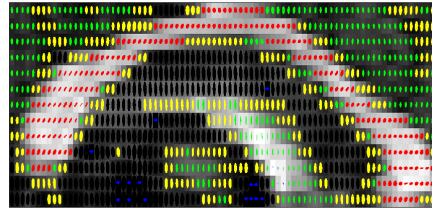
(b) Cholesky method



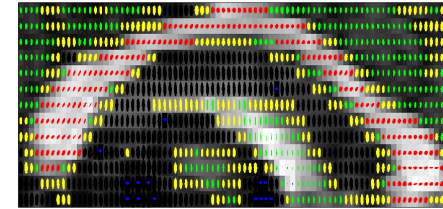
(c) Log Euclidean method



(d) Riemannian method



(e) Root Euclidean method



(f) Procrustes method

Figure 3.8: Segmentation of the CC from the midsagittal plane, superimposed over the FA map. The tensors in the CC are colored in red. All the methods distinguish between the CC and the fornix.

Segmentation method	Euclidean mean	Log Euclidean mean	Riemannian mean	Cholesky mean	Root Euclidean mean	Procrustes mean
Euclidean	2.4759e-07	2.6222e-07	2.6818e-07	2.5269e-07	2.5071e-07	2.5090e-07
Log Euclidean	0.9542	0.9012	0.9039	0.9070	0.9165	0.9123
Riemannian	1.0050	0.9616	0.9587	0.9677	0.9721	0.9704
Cholesky	.8803e-05	9.7007e-05	9.8081e-05	9.5853e-05	9.6275e-05	9.6047e-05
Root Euclidean	3.4259e-04	3.4433e-04	3.4662e-04	3.3904e-04	3.3748e-04	3.3772e-04
Procrustes	8.1587e-05	8.1527e-05	8.2411e-05	8.0192e-05	8.0063e-05	8.0003e-05

Table 3.3: The variance of the CC using Euclidean and non-Euclidean methods with different means.

in the CC, then the variance of the CC is calculated as follows:

$$\text{var}(\text{CC}) = \frac{1}{m(cc) - 1} \sum_{i \in C_{cc}} D(\mathbf{A}_i, \bar{\mathbf{A}}_{C_{cc}})^2, \quad (3.44)$$

where $D(.,.)$ and $\bar{\mathbf{A}}_{C_{cc}}$ is the, Euclidean or non-Euclidean, distance and mean of the tensors in the CC respectively. The variance of the segmented CC using Euclidean and non-Euclidean methods with different means (Equation 3.44) has been calculated in Table 3.3. The result confirmed that each method achieves the minimum variance with its mean as the main diagonal containing the minimum values (in blue).

How to choose the number of clusters

In a K -means algorithm, the cluster size is needed as an input to the algorithm. To choose the appropriate cluster size, silhouette method (Rousseeuw, 1987) is used. The K with the largest overall mean of the silhouette values, $\text{mean}(\text{sil})$, is the appropriate cluster size. Suppose we have a set of covariance matrices $\mathbf{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_M\}$ clustered into K clusters. The silhouette value of \mathbf{A}_i , where $i \in C_p$ and $p \in \{1, \dots, K\}$, is a measure of how similar \mathbf{A}_i is to the covariance matrices in cluster p , when compared to the covariance matrices in cluster j , where $j \in \{1, \dots, K\}$ and $j \neq p$. The silhouette value of \mathbf{A}_i is calculated as follows:

$$\text{sil}(\mathbf{A}_i) = \frac{b(\mathbf{A}_i) - a(\mathbf{A}_i)}{\max\{a(\mathbf{A}_i), b(\mathbf{A}_i)\}}.$$

where $a(\mathbf{A}_i)$ is the mean distance between the covariance matrix \mathbf{A}_i and other covariance matrices in cluster p . Let \bar{d}_j represent the mean distance between \mathbf{A}_i and the covariance matrices in cluster j . Then $b(\mathbf{A}_i) = \min_{j \neq p} \bar{d}_j$. The mean of the silhouette values is $mean(sil) = \sum_{i=1}^M sil(\mathbf{A}_i)/M$. The ROI is clustered into 2 to 8 clusters (as a small suitable number of clusters is more preferable than a large number of clusters to reduce clustering time) using Hartigan's method. The CC is not visible at 2 and 3 clusters, and hence the appropriate cluster size for segmentation of the CC needs to be 4 or more. Then $mean(sil)$ is calculated for 4, \dots , 8 clusters using Euclidean and non-Euclidean methods (see Table 3.4). For all non-Euclidean methods, the largest value of $mean(sil)$ is obtained at 5 clusters (see Table. 3.4), and hence for the segmentation of the CC the most appropriate cluster size is 5. For the Euclidean method, the $mean(sil)$ for 4 clusters is slightly larger than 5 clusters. However, to be able to compare the segmentation results using the different methods, we choose 5 clusters for the Euclidean method too.

K	Eucl.	Cholesky	log Eucl.	Riemannian	Root Eucl.	Procrustes
4	0.574	0.520	0.628	0.630	0.559	0.565
5	0.566	0.567	0.668	0.668	0.603	0.613
6	0.524	0.472	0.543	0.542	0.537	0.546
7	0.538	0.500	0.575	0.573	0.557	0.564
8	0.512	0.443	0.479	0.477	0.498	0.512

Table 3.4: The $mean(sil)$ is calculated for 4 to 8 clusters. We start with 4 since the CC is distinguishable from 4 clusters upwards. The highest average of silhouette values is at 5 clusters for non-Euclidean methods.

An alternative way to determine the appropriate cluster size is using the variance of the CC. We suppose that the tensors in the CC are homogeneous and hence we seek a small variance of the CC, with a small K to save clustering time. High variance of the CC indicates that the tensors in the CC are not homogenous, which could happen if some of the tensors that close to the boundary of the CC are included in the segmentation. Since with 2 and 3 clusters the CC is not clearly distinguishable, the variance of the CC region (using Equation 3.44) at cluster size $K \in \{4, \dots, 8\}$ is presented in Table 3.5. In fact, the size of the CC decreases with

increasing number of clusters. However, the variances of the CC are similar from cluster size 5 upwards, whilst there is a larger reduction in variance values from cluster size 4 to 5. For this reason, we suggest that 5 is the most appropriate choice for segmentation of the CC.

Method	4	5	6	7	8
Euclidean	2.847e-07	2.476e-07	2.408e-07	2.272e-07	2.259e-07
Log Eucl.	3.109	0.901	0.895	0.896	0.897
Riemannian	3.185	0.959	0.953	0.956	0.955
Cholesky	2.570e-04	9.585e-05	9.099e-05	9.097e-05	9.068e-05
Root Eucl.	9.630e-04	3.375e-04	3.245e-04	3.245e-04	3.223e-04
Procrustes	2.165e-04	5.713e-05	5.308e-05	5.193e-05	5.121e-05

Table 3.5: The variance of the CC using 5 clusters is significantly smaller than the variance of the CC using 4 clusters. However the variances are similar from 5 clusters upwards.

Evaluation of the segmentation

To evaluate the segmentation results obtained by the different methods, a manual segmentation is performed by an expert (see Figure 3.9) and the accuracy and specificity (Xu, Mandal, Long, Cheng, & Basu, 2012) are calculated (Figure 3.10) as follows:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (3.45)$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (3.46)$$

where TP refers to true positive (the number of tensors in the CC that are segmented correctly), TN refers to true negative (the number of tensors in the background that are segmented correctly), FP refers to false positive (the number of tensors in the CC that are incorrectly segmented as part of the background) and FN refers to false negative (the number of tensors in the background that are incorrectly segmented as part of the CC). The log Euclidean and Riemannian methods provide the highest accuracy and specificity of the segmentation and hence they outperform

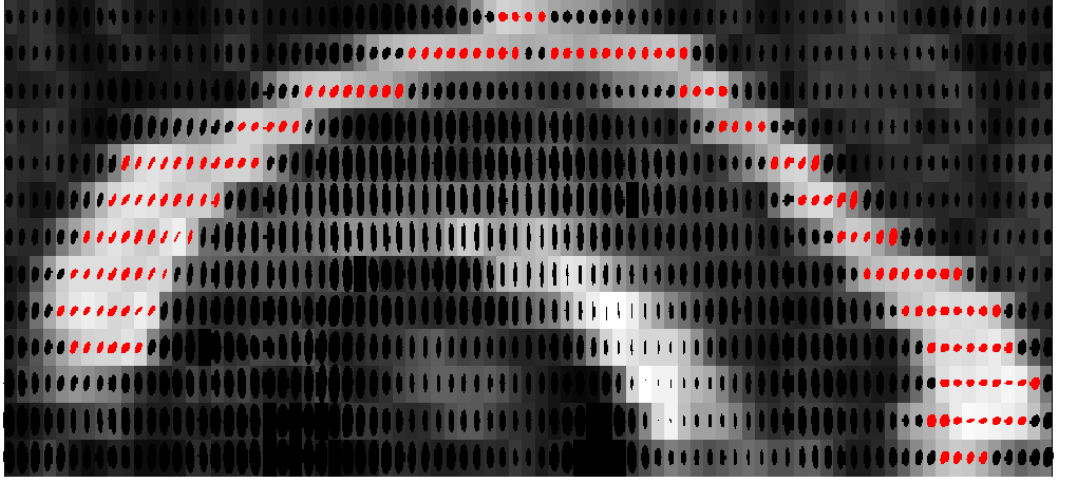


Figure 3.9: The manual segmentation of the ROI. The tensors in the CC are colored red while other tensors are colored black.

other methods for the segmentation of the CC. The Euclidean method is the least accurate method for the segmentation.

To evaluate the size and shape of the tensors in the CC, we used the standard error of the mean (SEMean). The size and shape of a tensor \mathbf{A} can be measured using the determinant of \mathbf{A} , $|\mathbf{A}|$, and $\text{FA}(\mathbf{A})$ respectively. The most accurate method for estimating the size and shape of the tensors in the CC is the method that provides the least SEmean. Suppose sd is the standard deviation, then $\text{SEMean} = sd/\sqrt{n}$, where n is the number of the tensors in the CC. The results are shown in Figure 3.11. It is clear that the Euclidean method produces the highest values of SEmean of FA and $|\mathbf{A}|$ which means the highest variation of shape and size of the tensors in the CC. High variation of size and shape of the tensors in the CC are not expected in the healthy brain. The log Euclidean and Riemannian methods provide the smallest variation of FA and $|\mathbf{A}|$ and hence they are preferable for the segmentation of the CC.

3.5.3 Three Dimensional Segmentation Using Hartigan's Method

A ROI is selected from the middle of the brain to perform the 3D segmentation of the CC. Similar to the 2D segmentation, at 2 and 3 cluster sizes the CC is not yet

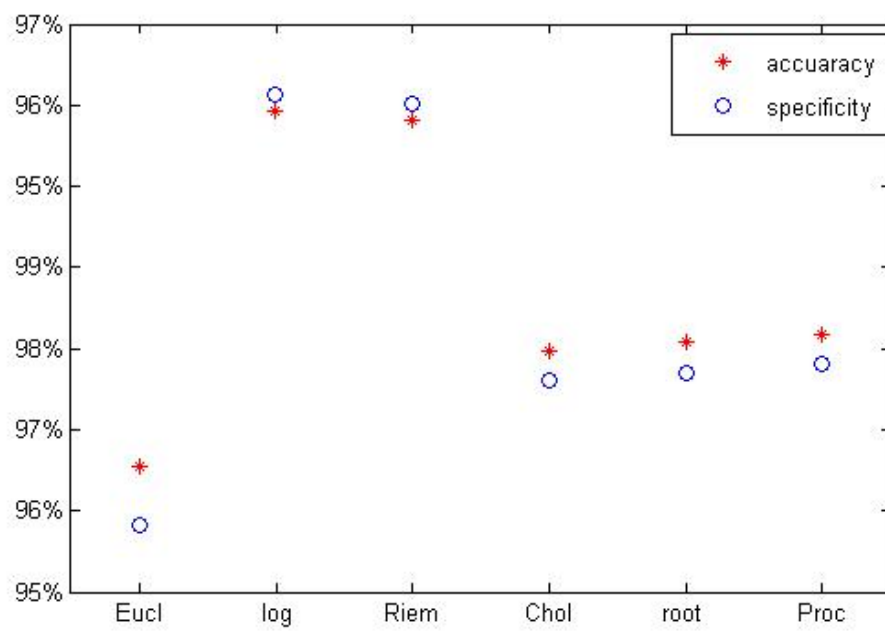


Figure 3.10: The accuracy and specificity of the segmentation using the Euclidean, log Euclidean, Riemannian, Cholesky, root Euclidean and Procrustes methods. Log Euclidean and Riemannian provide the highest accuracy and specificity.

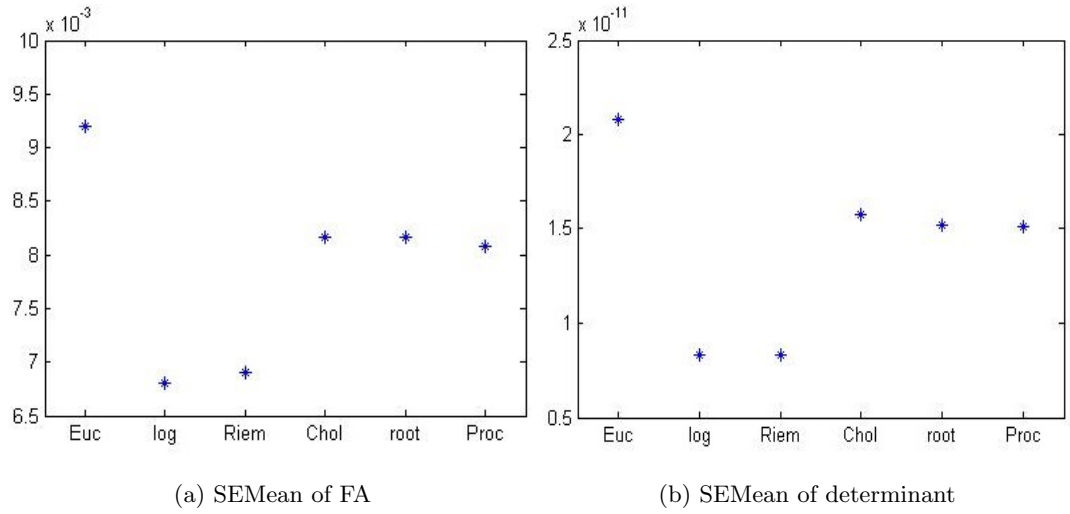


Figure 3.11: The SEMean of FA and determinant (det) for the CC region. The least SEMean values are obtained by the log Euclidean and Riemannian methods.

separated as a cluster from other regions in the ROI. This is the case for all the methods. The CC is clearly distinguishable as one cluster at 4 and 5 clusters but there is some noise around the CC (i.e. disjoint components that clustered with the CC as one cluster but they are not part of the CC). In fact there is more noise around the CC at cluster size 4 than 5. The CC is no longer represented by one cluster when the cluster size is increased to 6 or more. Therefore, the appropriate cluster size for segmentation of the CC is 5. The segmentation results after removing the noise around the CC, superimposed over FA background of an axial slice, is shown in Figure 3.12. The log Euclidean and Riemannian are the only methods that yield a full splenium's shape (see the yellow arrows in Figures 3.12c and 3.12d).

Descriptive statistics of the three dimensional segmentation

To compare the FA and the size of the tensors in the 3D CC obtained using the different methods, we draw box plots of the FA (Figure 3.13) and size of tensors in the CC (in Figure 3.14). The log Euclidean and Riemannian methods produce the smallest variation of FA and size of the tensors. In addition, they yield the highest average of FA and the smallest average of size of the tensors. This indicates that the segmentation using the log Euclidean and Riemannian methods yield a CC with

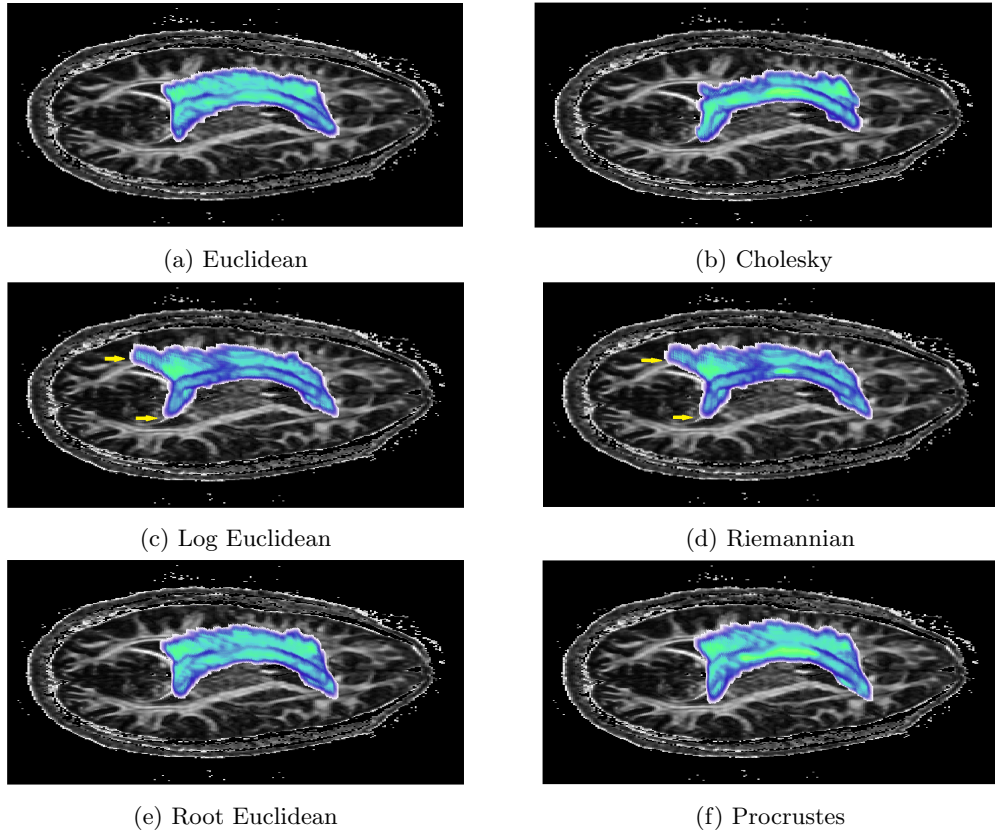


Figure 3.12: The three dimensional segmentation of the CC after removing the small objects around the CC. The log Euclidean and Riemannian methods produce a full shape of the splenium of the CC (the yellow arrows).

highly directional and small-sized tensors. The average of the FA and the size values using the Euclidean, Cholesky, root Euclidean and Procrustes methods are similar. The Euclidean method yield the highest variation of size of the tensors with the largest number of outliers in FA and size (i.e. tensors with small FA and with large size), shown by the stars down the lower whisker (in Figure 3.13) and up the upper whisker (in Figure 3.14).

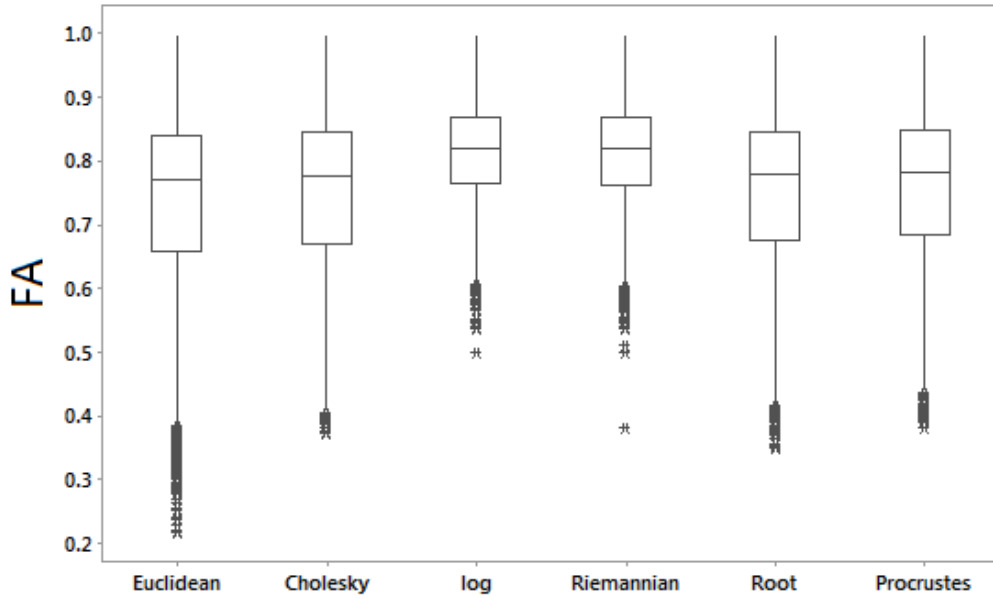


Figure 3.13: The box plot of the FA values of the tensors in the CC. The log Euclidean and Riemannian methods yield the highest average and the smallest variation of FA of the tensors.

3.6 Discussion and Conclusions

We generalized Hartigan’s method to work with the f -mean, so that we could use it to cluster diffusion tensor fields. For 2D and 3D segmentation of the CC, the log Euclidean and Riemannian methods provide more accurate segmentation than the other methods (Euclidean, Cholesky, root Euclidean and Procrustes). The log

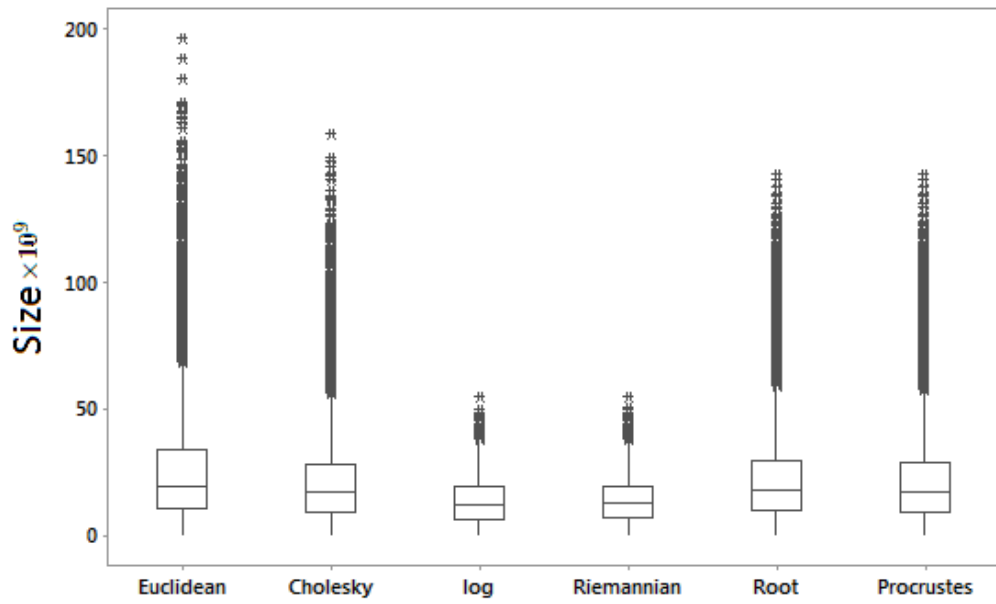


Figure 3.14: The box plot of size ($\times 10^9$) of the tensors in the CC. The log Euclidean and Riemannian methods yield the smallest average and the smallest variation of size of the tensors.

Euclidean and Riemannian methods provide less variations in size and shape of the tensors in the CC. As the computation of the log Euclidean mean is much faster than the Riemannian mean (Arsigny et al., 2007), the log Euclidean method is more efficient. Generally, the computation of the Euclidean distance is faster than the computation of the non-Euclidean distances. In addition, the Riemannian and Procrustes means are computationally expensive compared with the other means (Euclidean, Cholesky, root Euclidean and log Euclidean), as they need iterative numerical methods for the computations (Zhou et al., 2016).

Using hierarchical clustering (single link), the segmentation of the CC using the Riemannian method provide the smallest average number of holes in the CC and hence it is preferred over the other methods. Generally, hierarchical segmentation (single link) using Euclidean and non-Euclidean methods produce holes on the CC whereas no holes are produced when using Hartigan’s method. Therefore Hartigan’s method outperforms the hierarchical segmentation (single link).

Chapter 4

Modeling Diffusion Directions of the CC

Diffusion Tensor Imaging (DTI) has been used to study the characteristics of Multiple sclerosis (MS) in the brain. The von Mises-Fisher distribution (vmf) is a probability distribution for modeling directional data on the unit hypersphere. In this chapter, the diffusion directions of the normal appearing Corpus Callosum (CC) are modeled as a mixture of vmf distributions for MS and healthy subjects. This allows us to determine the distinct mean diffusion directions in the normal-appearing CC. Also it helps to cluster the CC depending on the diffusion directionality. Using Bayesian information criterion (BIC), we find that there are at least three distinct mean directions in the three dimensional CC. Higher diffusion concentration around the mean directions and smaller sum of angles between the mean directions are observed on the normal-appearing CC of the MS subjects as compared to the healthy subjects. The maximum angle between the three mean directions is proposed as a curvature threshold to provide stopping criteria for fiber tracking in the CC.

This chapter is introduced in Section 4.1. Then background and related work of DTI in MS and vmf distribution are reviewed in Sections 4.2 and 4.3. Diffusion directions in the CC are modeled and the results are presented in Sections 4.4 and 4.5. A brief summary and discussion are provided in Section 4.6.

4.1 Introduction

Multiple Sclerosis (MS) is an immune-mediated neurological disease. The Corpus Callosum (CC) is the largest bundle of commissural fibers in the brain, and damage to it has been associated with increased risk of developing disability in MS. Microscopic damage due to MS is known to occur also outside of macroscopic lesions, and DTI was found to be sensitive to such abnormalities (Werring et al., 1999; Bammer et al., 2000; Guo, MacFall, & Provenzale, 2002; Rovaris et al., 2002; Ciccarelli et al., 2003; Oh, Henry, Genain, Nelson, & Pelletier, 2004; Ge et al., 2004; Hesseltine et al., 2006). Werring et al. (1999) showed that FA in normal appearing white matter in MS subjects are lower as compared with healthy controls. In the MS subjects, they demonstrated that the FA values in different types of lesions (either enhancing or non-enhancing) are lower ($p - value < 0.001$) than the FA values in normal appearing white matter. The average FA values (plus standard deviation) on lesions are less than 0.53 (Werring et al., 1999).

The von Mises-Fisher distribution (vmf) is one of the most basic probability distributions for modeling directional data on the unit hypersphere (Mardia & Jupp, 2009). The use of a finite mixture of vmf distributions to cluster directional data on the unit sphere has been proposed by Banerjee et al. (2005). They used an Expectation-Maximization (EM) algorithm to estimate the parameters of the mixture model. For the diffusion imaging, the vmf distribution has been used in the literature mainly to model the orientation distribution function (odf). For example, McGraw et al. (2006) model the odf of High Angular Resolution Diffusion Imaging (HARDI) as a mixture of vmf distributions. They also use this model for segmentation using synthetic and real HARDI data. In Bhalerao and Westin (2007), they proposed a $5D$ hyper spherical model for HARDI data using a mixture of vmf distributions. A method for reconstructing Diffusion Weighted signal using a continuous mixture of vmf distributions has been introduced in Kumar et al. (2008). They validate the method using synthetic and real brain data. A probabilistic fiber tracking algorithm using a particle filtering technique and vmf sampling has been

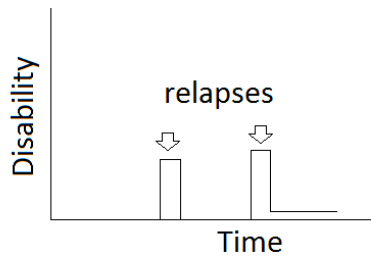
proposed in Zhang, Hancock, Goodlett, and Gerig (2009). In Painter and Hillen (2013), they build a model of glioma growth using Diffusion Tensor Imaging (DTI) and a bimodal vmf distribution. The vmf distribution, spherical harmonic (SH) expansion and Fractional anisotropy (FA) have been used in Reynolds et al. (2014) for the classification of Alzheimer’s disease. Comparing the results, they found that vmf and SH outperform the FA in the classification.

4.2 Background and Related Work

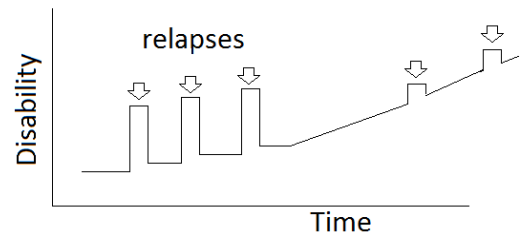
In this section, the main types of MS are discussed. Then related work of DTI measures to distinguish between healthy and MS subjects is studied.

4.2.1 Multiple Sclerosis

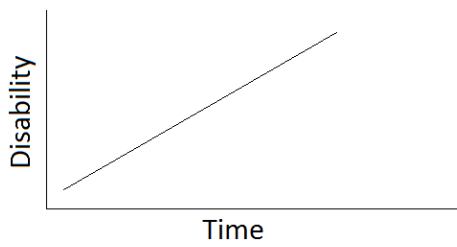
Mainly there are four types of MS: relapsing-remitting, primary progressive, secondary progressive and progressive-relapsing (see Figure 4.1). The most common type is relapsing-remitting MS (Figure 4.1a) which is characterized by recurring inflammatory events, associated with demyelinating lesions within the white matter of the brain and spinal cord. The majority of patients after a variable period of relapsing-remitting course develop the so-called secondary-progressive form of MS, characterised by a chronic accumulation of disability, with less and less acute events (Figure 4.1b). People with primary progressive MS (Figure 4.1c) are characterised by a continuous worsening their condition. The primary progressive MS is not very common as around 10%–15% of people with MS are diagnosed with it (D. H. Miller & Leary, 2007). Progressive-relapsing is a rare type of MS, where the disability is progressing with relapses happening from time to time (Figure 4.1d). The main difference between relapsing-remitting and progressive-relapsing is that in the former there is no worsening in the disability between relapses. Apart from the main four types of MS, Benign MS is the type of MS when the disability is low for at least 15 years of disease duration (Amato et al., 2006; Ceccarelli et al., 2008).



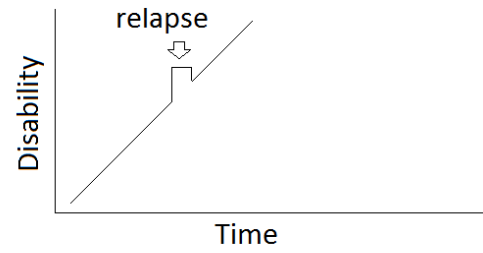
(a) Relapsing-remitting



(b) Secondary progressive



(c) Primary progressive



(d) Progressive-relapsing

Figure 4.1: The disability level varies with the time for each type of MS. No disability at the early stages of relapsing-remitting MS, but there is a slight and very gradual increase in disability level between replaces with the time. The disability level increases faster with the time in progressive-relapsing, primary progressive and secondary progressive MS.

4.2.2 DTI measures in MS

Generally the Fractional Anisotropy (FA) of the brain white matter (including the CC) has been reported to be decreased and the Mean Diffusivity (MD) to be increased for MS subjects (Werring et al., 1999; Tievsky, Ptak, & Farkas, 1999; Filippi et al., 2001; Coombs et al., 2004; Ge et al., 2004). Scheel et al. (2014) studied the relationship between the thickness of Retinal nerve fibre layer (RNFL) and FA values in the brain white matter for group of healthy and MS subjects. They found a positive correlation between the thickness of RNFL and FA values in the MS subjects whereas no such relationship exists in the healthy controls.

Rather than FA and MD, other measures such as axial diffusivity, radial diffusivity and relative anisotropy have been reported to indicate abnormalities due to MS. The axial diffusivity refers to the value of the principal eigenvalue of diffusion tensor, Radial diffusivity is the average of the second and third eigenvalues of diffusion tensor and Relative anisotropy is the standard deviation of eigenvalues divided by the mean of eigenvalues. For more details see Chapter 2. S.-K. Song et al. (2005) studied the affect of demyelination on radial diffusivity in mouse brains. They demonstrated that the radial diffusivity increases with demyelination and decreases with remyelination. Budde et al. (2008) studied the correlation between the axial diffusivity, radial diffusivity and relative anisotropy and the neurological disability in a mouse model of MS. Axial diffusivity was correlated negatively with the neurological disability while radial diffusivity and relative anisotropy did not show any association with the disability. Koenig et al. (2015) stated that both the axial diffusivity and radial diffusivity for MS subjects are significantly different as compared with healthy controls.

Abnormalities due to MS have also been suggested in the gray matter of the brain. For example, Tovar-Moll et al., 2009 studied the normal-appearing thalamus (a brain gray matter located just above the brain stem) and found that FA and MD are significantly higher in MS subjects as compared with healthy volunteers. Abnor-

malities in cortical gray matter was greater in Secondary Progressive MS subjects as compared with Relapsing-remitting MS subjects (Yaldizli et al., 2016).

4.3 Von Mises-Fisher Distribution

The vmf is a probability distribution on the $(d-1)$ dimensional sphere in \mathbb{R}^d (Mardia & Jupp, 2009). The vmf density function of the unit vector \mathbf{x} , given the mean direction $\boldsymbol{\mu}$ and the concentration around the mean k , is given by:

$$f(\mathbf{x}|\boldsymbol{\mu}, k) = c_d(k)e^{k\boldsymbol{\mu}^T\mathbf{x}}, \quad (4.1)$$

where $\|\boldsymbol{\mu}\| = 1$, $k > 0$ and $d \geq 2$. The normalizing constant $c_d(k)$ is given by:

$$c_d(k) = \frac{k^{d/2-1}}{(2\pi)^{d/2}I_{d/2-1}(k)}, \quad (4.2)$$

where $I_v(\cdot)$ is the modified Bessel function of the first kind at order v (Abramowitz & Stegun, 1964). If $d = 3$ the normalization constant can be written as

$$c_3(k) = \frac{k}{4\pi \sinh(k)}. \quad (4.3)$$

Intuitively, the concentration is the opposite of the variation and spread, as high concentration around the mean direction means small variation. This is illustrated in Figure 4.2 which shows three mean directions (in orange, green and blue). The data set in blue has the highest concentration.

The von Mises distribution is a special case of the vmf distribution, with $d = 2$, and is used to model circular data. The density function of the angle θ given the mean angle $\boldsymbol{\mu}$ and the concentration around the mean k is:

$$f(\theta|\boldsymbol{\mu}, k) = \frac{e^{k\cos(\theta-\boldsymbol{\mu})}}{2\pi I_0(k)} \quad (4.4)$$

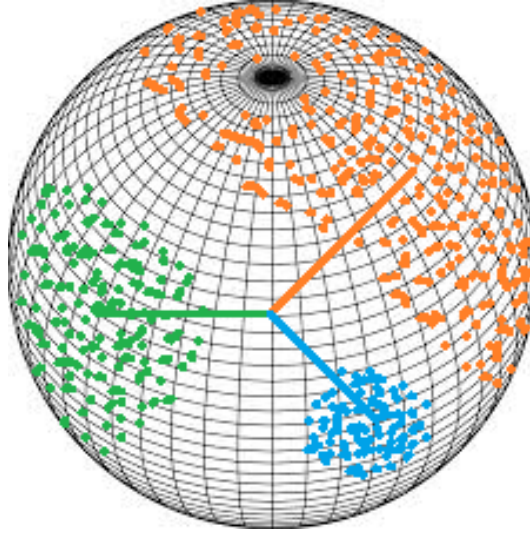


Figure 4.2: Three mean directions (in orange, green and blue). The data set in blue color has the highest concentration around the mean direction followed by the data set in green. The data set in orange has the smallest concentration around the mean direction.

where $I_0(k)$ is the modified Bessel function of order 0. This distribution with parameter μ and $1/k$ is the circular analogue to the normal distribution with parameters μ and σ^2 .

The antipodal symmetric property implies that the probability of each vector in a sphere equals to the probability of the negative vector as follows:

$$p(\mathbf{x}) = p(-\mathbf{x}). \quad (4.5)$$

The vmf distribution does not provide the antipodal symmetric property. Examples of antipodal symmetric distributions are Bingham (Bingham, 1974) and Watson (Watson, 1965) distributions. The diffusion in each tensor satisfies the antipodal symmetry property as the orientation of the diffusion is arbitrary (i.e. diffusion to left is equivalent to the diffusion to right, diffusion to superior is equivalent to the diffusion to inferior, diffusion to anterior is equivalent to the diffusion to posterior). Hence, Bingham and Watson distributions are suitable for modeling diffusion using orientation distribution function (odf), where each voxel in the brain is modeled as sphere. To be able to use the vmf distribution for modeling the odf,

an adaptation to the distribution is required. For example, to reconstruct Diffusion Weighted signal, Kumar et al. (2008) proposed the paired vmf distribution \mathcal{M} . The paired vmf distribution satisfies the antipodal symmetric property, by averaging $p(\mathbf{x})$ and $p(-\mathbf{x})$ as follows:

$$\mathcal{M}(\pm\mathbf{x}; \boldsymbol{\mu}, k) = \frac{c_d(k)e^{k\boldsymbol{\mu}^T\mathbf{x}} + c_d(k)e^{-k\boldsymbol{\mu}^T\mathbf{x}}}{2} = c_d(k)\cosh(k\boldsymbol{\mu}^T\mathbf{x}). \quad (4.6)$$

The antipodal symmetric property is only needed if the modeling is for each tensor. As we modeling all the tensors in the CC (and not each tensor), the antipodal symmetry is not relevant. Therefore, the vmf is the appropriate option for the modeling.

4.3.1 Maximum Likelihood Estimators

Estimating the parameters of vmf distribution can be performed using the maximum likelihood method which is based on maximizing the log-likelihood function (Myung, 2003). The parameter estimations are provided in Banerjee et al. (2005) and explained as follows.

The likelihood function of the parameters is obtained as follows:

$$P(X|\boldsymbol{\mu}, k) = \prod_{i=1}^n f(\mathbf{x}_i|\boldsymbol{\mu}, k), \quad (4.7)$$

where $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is an independent sample from the vmf distribution. Obtaining log of the likelihood function yields

$$\ln(P(X|\boldsymbol{\mu}, k)) = \sum_{i=1}^n \ln(f(\mathbf{x}_i|\boldsymbol{\mu}, k)) = n\ln(c_d(k)) + k\boldsymbol{\mu}^T \sum_{i=1}^n \mathbf{x}_i. \quad (4.8)$$

As there is a constraint that $\|\boldsymbol{\mu}\| = 1$, which implies that $\boldsymbol{\mu}'\boldsymbol{\mu} = 1$, the Lagrange multiplier λ (Rockafellar, 1993; Banerjee et al., 2005) can be used as follows:

$$\ln(P(X|\boldsymbol{\mu}, k)) = n \ln(c_d(k)) + k \boldsymbol{\mu}^T \sum_{i=1}^n \mathbf{x}_i + \lambda(1 - \boldsymbol{\mu}' \boldsymbol{\mu}). \quad (4.9)$$

To find the parameters, we need to differentiate Equation 4.9 with respect to $\boldsymbol{\mu}$, k and λ , then equate the equations to zero (Myung, 2003). We obtain the following equations:

$$\hat{k} \sum_{i=1}^n \mathbf{x}_i = 2\hat{\lambda} \hat{\boldsymbol{\mu}}, \quad (4.10)$$

$$-n \frac{d(c_d(\hat{k}))/dk}{c_d(\hat{k})} = \hat{\boldsymbol{\mu}}' \sum_{i=1}^n \mathbf{x}_i \quad (4.11)$$

and

$$\hat{\boldsymbol{\mu}}' \hat{\boldsymbol{\mu}} = 1. \quad (4.12)$$

Here, Equation 4.10 implies that

$$\hat{\boldsymbol{\mu}} = \frac{\hat{k} \sum_{i=1}^n \mathbf{x}_i}{2\hat{\lambda}}. \quad (4.13)$$

Substituting Equation 4.13 in Equation 4.12 yields

$$\hat{\lambda} = \frac{\hat{k}}{2} \left\| \sum_{i=1}^n \mathbf{x}_i \right\|. \quad (4.14)$$

Then substituting Equation 4.14 in Equation 4.13 gives the estimator of the mean direction $\hat{\boldsymbol{\mu}}$ as follows:

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^n \mathbf{x}_i}{\left\| \sum_{i=1}^n \mathbf{x}_i \right\|}. \quad (4.15)$$

To obtain the estimator of the concentration \hat{k} , we substitute the value of $\hat{\boldsymbol{\mu}}$ (Equation 4.15) in Equation 4.11 as follows:

$$- \frac{d(c_d(\hat{k}))/dk}{c_d(\hat{k})} = \frac{(\sum_{i=1}^n \mathbf{x}_i)' (\sum_{i=1}^n \mathbf{x}_i)}{n \left\| \sum_{i=1}^n \mathbf{x}_i \right\|} \quad (4.16)$$

Using the definition of $c_d(k)$ in Equation 4.2, it is easy to show that (Banerjee

et al., 2005)

$$-\frac{d(c_d(\hat{k}))/dk}{c_d(\hat{k})} = \frac{I_{d/2}(\hat{k})}{I_{d/2-1}(\hat{k})}. \quad (4.17)$$

Let $A(k) = \frac{I_{d/2}(k)}{I_{d/2-1}(k)}$, then Equation 4.16 implies

$$A(\hat{k}) = \frac{\|\sum_{i=1}^n \mathbf{x}_i\|}{n}. \quad (4.18)$$

The estimator \hat{k} can be obtained by solving Equation 4.18 numerically. Many approximations and solutions of \hat{k} have been proposed (Dhillon & Sra, 2003; Banerjee et al., 2005; Tanabe, Fukumizu, Oba, Takenouchi, & Ishii, 2007; H. Song, Liu, & Wang, 2012; Sra, 2012; Hornik & Grün, 2014b).

4.3.2 Mixture of vmf

Estimating the parameters of the mixture distribution of vmf, using the Expectation-Maximization (EM) algorithm, was studied by Dhillon and Sra (2003) and Banerjee et al. (2005). The EM algorithm (Dempster, Laird, & Rubin, 1977) use the complete-data log-likelihood (the log-likelihood on all the variables: known and unknown) to estimate the maximum likelihood estimates of the parameters. The density of a finite mixture H of vmf distributions (Banerjee et al., 2005) is given by

$$f(\mathbf{x}|\Theta) = \sum_{h=1}^H w_h f_h(\mathbf{x}|\theta_h), \quad (4.19)$$

where Θ denotes the sequence of $w_1, w_2, \dots, w_H, \theta_1, \theta_2, \dots, \theta_H$ are the parameters of the mixture density and $f_h(\mathbf{x}|\theta_h)$ is the density of the vmf distribution with parameter θ_h denotes the sequence of $\boldsymbol{\mu}_h, k_h$. Suppose we need to model the data set $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ as a mixture of vmf distribution. The data set X is the known (also called observed) variables. Suppose the corresponding set of unknown (also called hidden or latent) variables is $Z = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ which represents the clustering result of X into H groups. That is $\mathbf{z}_i \in \{1, \dots, H\}$ for $i \in \{1, \dots, n\}$, where $\mathbf{z}_i = h$ if \mathbf{x}_i is sampled from $f_h(\mathbf{x}|\theta_h)$ (Banerjee et al., 2005). The complete-data log-likelihood

function is given by

$$\ln(P(X, Z|\Theta)) = \sum_{i=1}^n \ln(w_{z_i} f_{z_i}(\mathbf{x}_i|\theta_{z_i})). \quad (4.20)$$

The EM algorithm iterates between two steps: Expectation and Maximization steps. In the expectation step, the expected complete-data log-likelihood is calculated using the conditional probability $p(h|\mathbf{x}_i, \Theta)$. That is because the complete-data log-likelihood is unknown as Z is unknown. Then the expected complete-data log-likelihood is maximized in the Maximization step. As in Dhillon and Sra (2003) and (Banerjee et al., 2005), maximizing the expectation in Equation 4.20 with the constraints $\boldsymbol{\mu}_h' \boldsymbol{\mu}_h = 1$ and $k_h \geq 0$ yields

$$w_h = \frac{\sum_{i=1}^n p(h|\mathbf{x}_i, \Theta)}{n} \quad (4.21)$$

$$\hat{\boldsymbol{\mu}}_h = \frac{\sum_{i=1}^n \mathbf{x}_i p(h|\mathbf{x}_i, \Theta)}{\|\sum_{i=1}^n \mathbf{x}_i p(h|\mathbf{x}_i, \Theta)\|} \quad (4.22)$$

$$A(\hat{k}_h) = \frac{\|\sum_{i=1}^n \mathbf{x}_i p(h|\mathbf{x}_i, \Theta)\|}{\sum_{i=1}^n p(h|\mathbf{x}_i, \Theta)}, \quad (4.23)$$

where $A(k_h) = \frac{I_{d/2}(k_h)}{I_{d/2-1}(k_h)}$. The equations 4.21, 4.22 and 4.23 are used in the Maximization step to estimate the parameters' values iteratively. The EM algorithm uses soft-assignment where each point is assigned a probability of being a member in each mixture component. When the soft-assignment algorithm converges, any point is assigned to exactly one cluster (i.e one mixture component). An alternative version is hard-assignment where the probability of each point is either 1 for member in the mixture component or 0 otherwise. This has been performed by adding a hardening step between the Expectation and maximization steps (Banerjee et al., 2005). Banerjee et al. (2005) demonstrated that soft-assignment algorithm outperforms hard-assignment algorithm. For more details see Dhillon and Sra (2003) and (Banerjee et al., 2005).

Initializing the EM algorithm can be performed in various ways e.g. randomly or using a k-mean algorithm such as spherical K -means (Dhillon & Modha,

2001). Generally, two stopping criteria for the EM algorithm can be implemented. The EM algorithm can stop if the number of iterations has exceeded the maximum number of iterations or if the algorithm converges. The algorithm converges if the relative change in the log-likelihood or the parameter value is less than a threshold (Hornik & Grün, 2014a). The relative change in log-likelihood refers to the absolute change in the log-likelihood value divided by the log-likelihood value in the previous iteration.

The Bayesian information criterion (BIC), also known as Schwarz criterion, is used for model selection to choose the most suitable model (Schwarz et al., 1978). Minimizing BIC corresponds to maximizing the log likelihood function and hence the model with the smallest BIC is the preferred model. The BIC is defined as:

$$BIC = g \log(n) - 2 \log(\hat{L}), \quad (4.24)$$

where g is the number of the parameters to be estimated, n is the sample size and \hat{L} is the maximized likelihood function of the model. To confirm the results from BIC, Akaike information criterion (AIC) is used. Similarly to BIC, the AIC is used for model selection (Akaike, 1973) and the model with the smallest AIC is the preferred model. The AIC is defined as:

$$AIC = 2g - 2 \log(\hat{L}). \quad (4.25)$$

4.4 Modeling Diffusion Directions of the CC

In this section, we modeled the principal eigenvectors of the diffusion tensors in the CC, of a group of healthy and MS subjects, as a mixture of vmf distributions to study the means and the concentrations of water diffusion in the CC. In this section, simulation study is provided first to evaluate the modeling then the imaging protocol of the brain images is presented and the methodology is explained.

4.4.1 Simulation Study

To evaluate the performance of the modeling, we used a simulation study of 1000 replication and a sample size $n = 746$. First fixed values for the parameters $\boldsymbol{\mu}$, k and w are given as follows:

$$\boldsymbol{\mu}_1 = (-0.9658, 0.2155, 0.1440, 0.9513),$$

$$\boldsymbol{\mu}_2 = (0.9513, -0.2235, 0.2123),$$

$$\boldsymbol{\mu}_3 = (-0.9385, -0.1180, 0.3244),$$

$$k = (15.9620, 11.0281, 14.5388),$$

$$w = (0.3579, 0.2426, 0.3995).$$

Second, a sample with the above parameters is generated from vmf distribution (Wood, 1994). Third, the parameters of the generated sample are estimated using the algorithm proposed in Banerjee et al. (2005). Fourth, to evaluate the estimation of k and w the relative errors are calculated (e.g. $RE(k) = |k - \hat{k}|/|k|$) and to evaluate the estimation of $\boldsymbol{\mu}$, $\boldsymbol{\mu}^T \hat{\boldsymbol{\mu}}$ is used ($\|\boldsymbol{\mu}\| = 1$ and hence $\boldsymbol{\mu}^T \hat{\boldsymbol{\mu}}$ should be close to 1). Last, the steps from second to fourth are replicated 999 times. The results of evaluation are shown in Table 4.1. The average values of RE(k) and RE(w) are small and the average value of $\boldsymbol{\mu}^T \hat{\boldsymbol{\mu}}$ is very close to 1. Therefore, we conclude the good performance of the algorithm proposed by Banerjee et al. (2005).

h	avg RE(k)	max RE(k)	avg RE(w)	max RE(w)	avg $\boldsymbol{\mu}^T \hat{\boldsymbol{\mu}}$	min $\boldsymbol{\mu}^T \hat{\boldsymbol{\mu}}$
1	0.0702	0.275	0.0719	0.206	0.999	0.933
2	0.0665	0.332	0.0629	0.232	0.999	0.922
3	0.0618	0.318	0.0533	0.226	0.999	0.996

Table 4.1: The average and maximum relative error (RE) of the parameters k and w over 1000 replications indicate the good estimation of the parameters k and w . The values of $\boldsymbol{\mu}^T \hat{\boldsymbol{\mu}}$ are close to 1 which indicate good estimation of $\boldsymbol{\mu}$.

4.4.2 Brain Images

The data was collected at the Neuroimaging Laboratory, Santa Lucia Foundation in Rome (Italy). The study was approved by Ethics Committee of Santa Lucia Foundation. Written informed consent was obtained from all subjects before entering the study. Diffusion-weighted imaging was obtained using a head-only 3.0 T scanner (Siemens Magnetom Allegra, Siemens Medical Solutions, Erlangen, Germany), using a twice-refocused spin echo echo-planar imaging (SE EPI) sequence (TR = 7000 ms, TE = 85 ms, maximum b factor = 1000 s/mm², isotropic resolution 2.3 mm³; matrix = 96×96; 60 slices), accomplished by collecting 7 images with no diffusion weighting ($b = 0$) and 61 images with diffusion gradients, applied in 61 non-collinear directions (scan time: 11 min).

The data consists of four (two males and two females) healthy subjects (the mean age \pm standard deviation is 62.80 ± 9.50) and nine (three males and six females) secondary progressive MS subjects (the mean age \pm standard deviation is 43.60 ± 10.07).

4.4.3 Methodology

The three dimensional CC is segmented using the Euclidean method (Chapter 3) as it is the fastest method for segmentation. The principal eigenvectors of the diffusion tensors in the CC represent the main diffusion directions in the CC. Hence, the unit vector \mathbf{x} represents the principal eigenvector, $\boldsymbol{\mu}$ is the mean diffusion direction and k is the concentration of the diffusion directions around $\boldsymbol{\mu}$. As the diffusion tensor is a 3×3 matrix, the principal eigenvector \mathbf{x} of the tensor is a 3×1 vector so $d = 3$. Substituting Equation 4.1 and 4.3 into Equation 4.19 gives:

$$f(\mathbf{x}|\Theta) = \frac{1}{4\pi} \sum_{h=1}^H \left[\frac{w_h k_h}{\sinh(k_h)} e^{k_h \boldsymbol{\mu}_h^T \mathbf{x}} \right]. \quad (4.26)$$

Fitting the mixture of vmf distributions using the EM algorithm (soft assignment) is performed using R (Hornik & Grün, 2014a). The initialization of EM

algorithm is based on a simple K -means. The algorithm stops if the relative change in the log-likelihood value is less than the threshold $2^{-26} = 1.490116e^{-8}$. The concentration parameter is calculated using the method proposed by Hornik and Grün (2014b). For implementation of the EM algorithm using R language see Hornik and Grün (2014a). The BIC values are used to evaluate the appropriate number of the mixture distributions.

The concentrations around the mean directions are compared between the healthy and the MS subjects. To compare the diffusion directions between the MS and healthy subjects, we calculate the angles between each two of the three mean directions (i.e. the angle between $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_2$, $\hat{\boldsymbol{\mu}}_2$ and $\hat{\boldsymbol{\mu}}_3$ and $\hat{\boldsymbol{\mu}}_3$ and $\hat{\boldsymbol{\mu}}_1$). For example, let ϕ be the angle between two mean directions $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_2$. As $\|\hat{\boldsymbol{\mu}}_i\| = 1$, for $i \in \{1, 2, 3\}$, then ϕ can be calculated as

$$\phi = \arccos(\hat{\boldsymbol{\mu}}_1 \cdot \hat{\boldsymbol{\mu}}_2) \quad (4.27)$$

As the diffusion is a symmetric process (we cannot distinguish between left to right diffusion and right to left which means the angle 0 is equivalent to the angle π), the angle between $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_2$ is calculated as the minimum of ϕ and $(\pi - \phi)$ and thus we get angles between 0 and $\pi/2$ for all the subjects. The mean FA values and variance of FA values of the CC are also compared between the healthy and the MS subjects.

4.5 Results

The BIC values showed that a mixture of vmf distributions for the CC direction is preferred over a single vmf distribution for both healthy and MS patients (see Figure 4.3 and 4.4). From the figures, it is clear that the BIC values for three mixture distributions are lower than one and two mixture distributions in all cases for both the MS and healthy subjects. Thus, at least three mixtures of vmf distributions are preferred, to model the diffusion in the whole CC. The BIC values for four and five

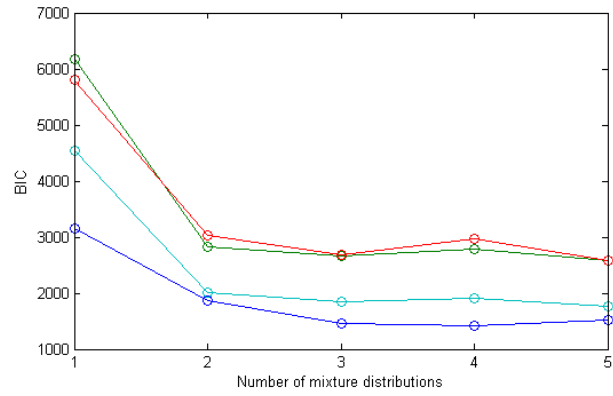


Figure 4.3: Each line represents a subject. For the four healthy subjects, Bayesian information criterion (BIC) for three mixture distributions are lower than one and two mixture distributions.

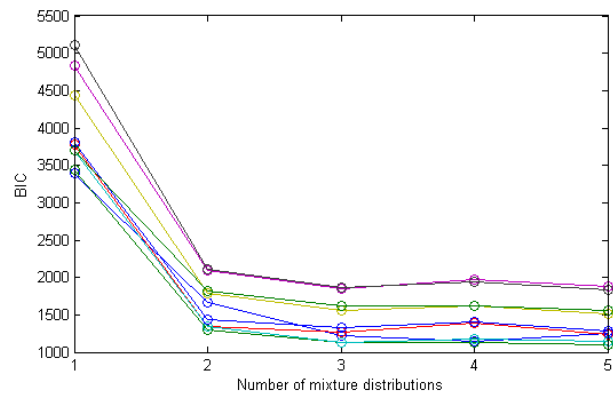


Figure 4.4: Each line represents a subject. For the nine MS subjects, BIC for three mixture distributions are lower than one and two mixture distributions.

Subjects	method	1	2	3	4	5
H1	BIC	3157.136	1860.041	1458.603	1416.096	1520.12
	AIC	3143.292	1827.738	1407.841	1348.645	1421.164
H2	BIC	4542.21	2006.344	1838.4	1910.053	1766.039
	AIC	4527.547	1972.131	1785.518	1836.74	1673.176
H3	BIC	5802.384	3032.325	2685.808	2959.942	2587.011
	AIC	5786.967	2996.353	2629.28	2882.859	2489.372
H4	BIC	6178.882	2819.302	2656.869	2759.882	2584.706
	AIC	6163.346	2783.053	2599.905	2703.698	2489.67

Table 4.2: The BIC and AIC values for single and two to five mixture distributions for the healthy subjects. The BIC and AIC values are similar and both of them indicate that three mixture distributions are preferred over two or single distribution for all the healthy subjects. Neither BIC nor AIC are consistent (i.e. monotonic) after three.

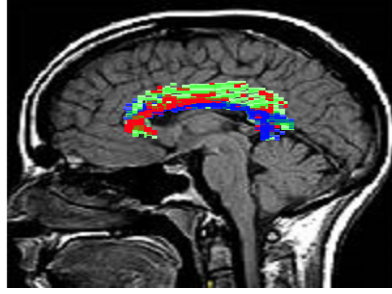
Subjects	method	1	2	3	4	5
MS1	BIC	3435.56	1292.075	1124.901	1131.547	1105.329
	AIC	3421.744	1259.8378	1074.2423	1081.5975	997.4169
MS2	BIC	3389.826	1659.518	1227.865	1149.057	1262.315
	AIC	3375.7461	1626.6648	1176.2389	1077.6881	1161.3409
MS3	BIC	3694.482	1343.147	1122.947	1178.456	1139.668
	AIC	3680.4428	1310.3901	1071.4716	1108.2624	1047.6798
MS4	BIC	3778.312	1340.747	1270.568	1389.037	1244.611
	AIC	3764.304	1308.061	1218.657	1318.994	1153.751
MS5	BIC	4830.463	2093.654	1853.73	1984.798	1886.546
	AIC	4815.654	2059.099	1799.43	1902.609	1792.193
MS6	BIC	4443.874	1791.948	1559.561	1671.535	1510.937
	AIC	4429.356	1758.071	1506.327	1561.708	1416.418
MS7	BIC	5106.718	2111.368	1861.17	1939.74	1835.256
	AIC	5091.731	2076.397	1806.216	1864.803	1744.872
MS8	BIC	3807.893	1437.16	1336.604	1399.068	1290.539
	AIC	3793.847	1404.385	1285.101	1355.087	1202.024
MS9	BIC	3699.169	1825.547	1624.997	1617.83	1567.906
	AIC	3685.141	1792.817	1573.563	1547.693	1477.07

Table 4.3: The BIC and AIC values for single and two to five mixture distributions for the MS subjects. The AIC values are similar to BIC values and both of them indicate that three mixture distributions are preferred over two or single distribution for all the MS subjects. Neither BIC nor AIC are consistent after three.

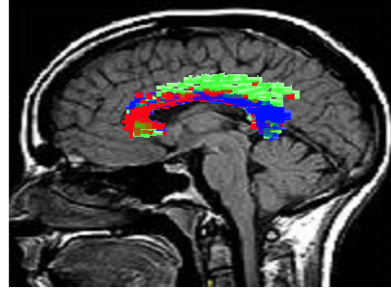
mixture distributions are either bigger or slightly smaller than BIC values for three mixture distributions. Hence, to be able to compare the models for all the subjects with the same number of mixture components, we choose to model the data using three mixtures of vmf distributions (i.e. $H = 3$ in Equation 4.26). Additionally, the AIC values are calculated (Tables 4.2 and 4.3) for the healthy and the MS subjects to confirm the BIC results. The AIC values are similar to the BIC values and hence this confirms the above result (i.e. that three mixture distributions are more appropriate to model and hence to compare the data). The mixture distributions of the CC for one healthy subject and one MS subject are shown in Figure 4.5.

Using Mann-Whitney test, we find that the sum of the angles between each two of the three mean directions are significantly higher for the healthy subjects than the MS subjects (p-value = 0.007). This result is shown in Figure 4.6a. Then we compare the concentration values between the MS and the healthy subjects (Figure 4.6b). The diffusion concentrations around the mean directions of the MS subjects are significantly higher than the diffusion concentrations of the healthy subjects (p-value = 0.001) using Mann-Whitney test. For the results of all the subjects see Tables 4.4 and 4.5. The tables show the mean directions, concentration, probability of the mixture components and the angles (ordered as: the angle between $\hat{\mu}_1$ and $\hat{\mu}_2$, $\hat{\mu}_2$ and $\hat{\mu}_3$, and $\hat{\mu}_3$ and $\hat{\mu}_1$).

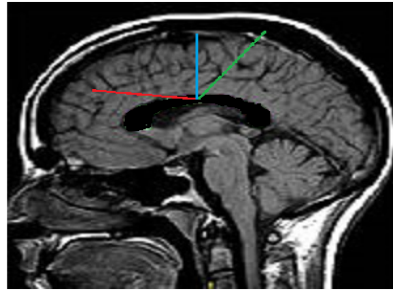
Using Mann-Whitney test, the difference in average of FA values in the CC between the healthy and the MS subjects is not significant (p-value= 0.604), see Figure 4.7a. However, using Mann-Whitney test, there is a significant difference in the variance of FA (p-value= 0.000) in the CC between the MS and the healthy subjects (Figure 4.7b).



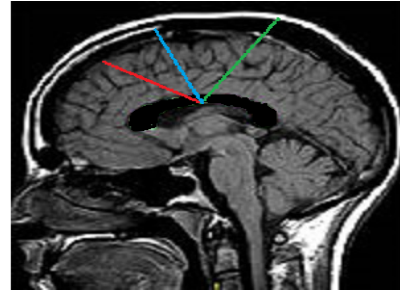
(a) The mixture directions for a MS subject.



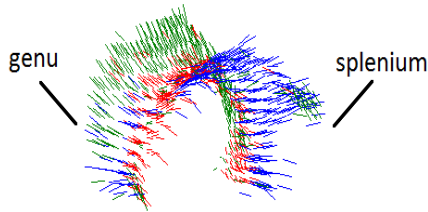
(b) The mixture directions for a healthy subject.



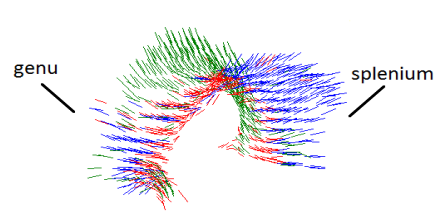
(c) The mean directions for the three mixture distributions in (a). The sum of angles between each two of the mean directions is 77.99.



(d) The mean vectors for the three mixture distributions in (b). The sum of angles between each two of the mean directions is 83.97.

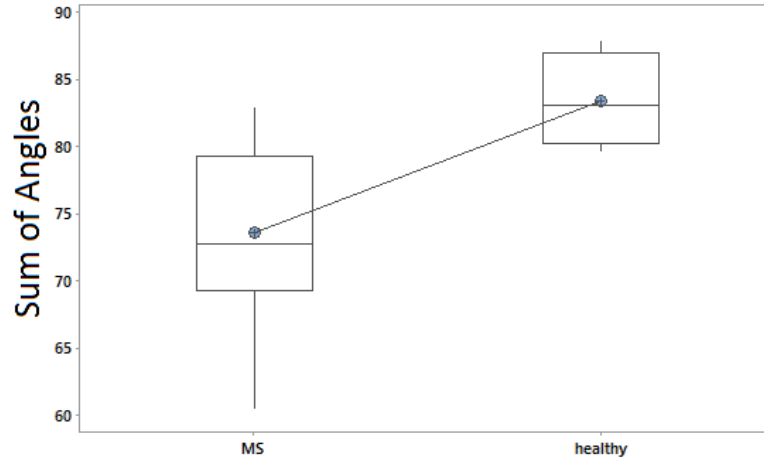


(e) The 3D view of the mixture directions in (a). The concentration values for the regions (blue, green, red) are (23.36, 13.48, 20).

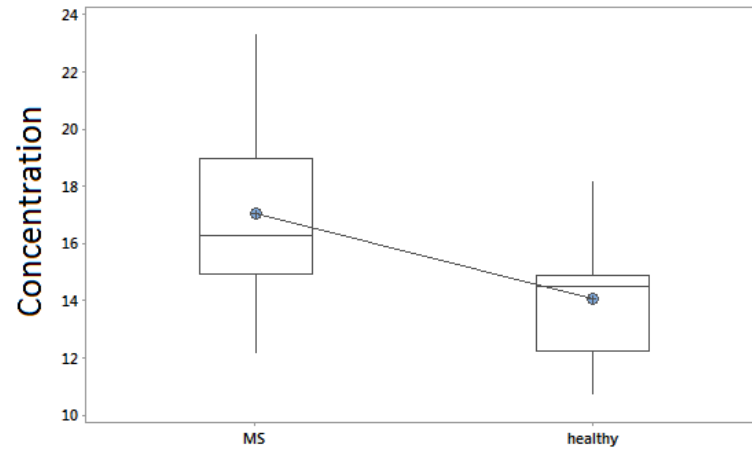


(f) The 3D view of the mixture directions in (b). The concentration values for the regions (blue, green, red) are (18.21, 12.02, 14.29).

Figure 4.5: The CC is clustered into three regions (blue, green and red) using the mixture of vmf distributions for a MS subject and a healthy subject. The sum of angles between each of the three mean directions are smaller for the MS subject. For both the MS subject and healthy subject, the diffusion directions on the blue and red regions are more concentrated than on the green region. The diffusion concentrations around the mean directions for MS subject are higher than for the healthy subject.

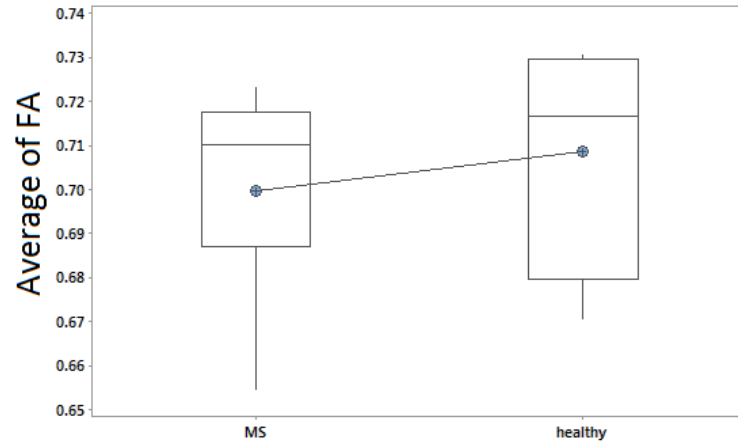


(a) Sum of the angles between each two of the three mean directions for the MS subjects are significantly smaller than for the healthy subjects.

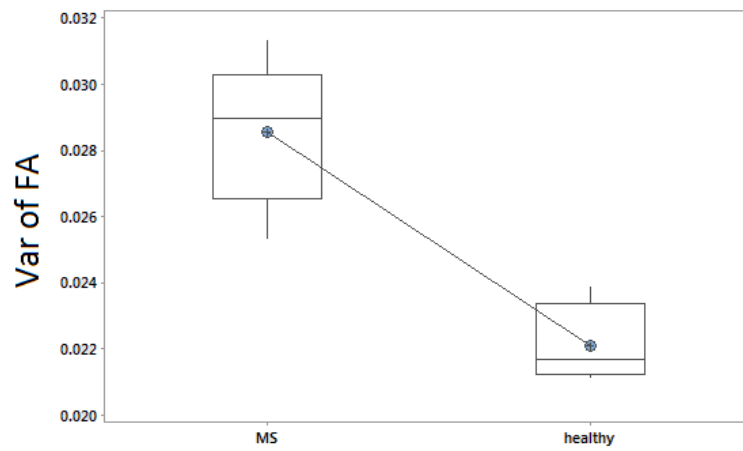


(b) Diffusion concentrations around the mean directions for the MS subjects are significantly higher than for the healthy subjects.

Figure 4.6: Box plots of sums of angles between the three mean directions and concentrations around the mean directions for the MS and the healthy subjects.



(a) Box plots of averages of FA. Insufficient evidence to suggest a difference in average of FA between the healthy and the MS subjects.



(b) Box plots of variances of FA. Sufficient evidence to suggest a difference in variance of FA between the healthy and the MS subjects.

Figure 4.7: Box plots of averages and variances of FA of the CC for the MS and the healthy subjects.

Subjects	$\hat{\mu}$	\hat{k}	Prob	Angle
H1 (n=746)	(-0.9658,0.2155,0.1440)	15.9620	0.3579	20.5443
	(0.9513,-0.2235,0.2123)	11.0281	0.2426	37.0965
	(-0.9385,-0.1180,0.3244)	14.5388	0.3995	21.9127
H2 (n=980)	(-0.9192,-0.3119,0.2403)	18.2062	0.3388	40.2720
	(0.9434,-0.0879,0.3197)	12.0247	0.3510	27.0855
	(-0.9907,-0.0579,0.1229)	14.2946	0.3102	16.6079
H3 (n=1260)	(-0.8906,-0.3706,0.2637)	14.7608	0.3619	32.6918
	(0.9406,0.2039,0.2716)	10.7144	0.3198	26.3603
	(-0.9908,-0.0038,0.1351)	14.4916	0.3183	23.1540
H4 (n=1311)	(-0.9853,0.1708,0.0034)	14.9478	0.3204	24.5238
	(0.9033,-0.1239,0.4107)	12.9591	0.3699	41.6507
	(-0.9599,-0.1045,0.2603)	14.7601	0.3097	21.7523

Table 4.4: The estimated values of the parameters of the vmf mixture model for the four healthy subjects i.e. mean directions ($\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\mu}_3$), diffusion concentrations \hat{k} around the mean directions and probabilities of the mixture components. The angles (last column) between the mean directions are ordered as the angle between $\hat{\mu}_1$ and $\hat{\mu}_2$, $\hat{\mu}_2$ and $\hat{\mu}_3$ then $\hat{\mu}_3$ and $\hat{\mu}_1$.

4.6 Conclusion and Discussion

We have used a mixture of vmf distributions to model the diffusion in the three dimensional CC. There are at least three different mean diffusion directions in the CC. For the MS subjects, the CC used in the modeling is the normal-appearing CC and the lesions are not included. That is because the directional information in the lesions is destroyed and hence including the lesions in the modeling might yield bias in the modeling results. Future work can address how the results might be affected if the regions with MS lesions in the CC are included in the analysis. On the normal-appearing CC of MS subjects the sum of angles between the mean diffusion directions are smaller with higher concentration around the mean directions compared with the healthy subjects. We hypothesize that this may be due to the atrophy in the CC of the MS subjects. The difference in the average of FA between the MS and the healthy subjects is not significant. This is due to the differences in the age between the MS and the healthy subjects, because the FA decreases with the age and hence both the MS and the healthy subjects have decreased FA. However, the variance of FA is significantly different between the healthy and the MS subjects

Subjects	$\hat{\boldsymbol{\mu}}$	\hat{k}	Prob	Angle
MS1 (n=739)	(-0.9998,-0.0171,0.0068)	15.9772	0.3207	15.9419
	(-0.9574,-0.1574,0.2422)	21.7699	0.3478	34.0371
	(0.9476,-0.0075,0.3193)	15.6433	0.3315	19.0653
MS2 (n=807)	(0.9623,-0.0404,0.2690)	15.6029	0.2404	22.8668
	(-0.9906,0.0518,0.1263)	15.2530	0.3457	14.3463
	(-0.9544,-0.1616,0.2510)	18.4241	0.4139	32.3928
MS3 (n=796)	(-0.9718,0.2144,0.0979)	21.8643	0.3505	23.1866
	(0.9350,-0.1868,0.3014)	17.9300	0.3229	29.0671
	(-0.9872,-0.1409,0.0752)	18.6192	0.3266	20.5299
MS4 (n=788)	(-0.9928,0.0431,0.1122)	16.9769	0.3604	8.2273
	(-0.9939,-0.0913,0.0618)	14.1895	0.2652	25.4873
	(0.9360,-0.0681,0.3455)	16.6264	0.3744	26.7036
MS5 (n=1029)	(-0.9664,-0.2466,0.0721)	18.9351	0.3460	17.9462
	(-0.9886,0.0572,0.1391)	14.9425	0.3129	26.3777
	(0.9502,0.0027,0.3116)	12.6465	0.3411	26.2978
MS6 (n=934)	(0.9495,-0.1283,0.2864)	13.4815	0.3597	34.4170
	(-0.9524,-0.2584,0.1614)	23.3600	0.3437	25.6290
	(-0.9867, 0.1611,0.0215)	20.0011	0.2966	17.9439
MS7 (n=1092)	(-0.9799, 0.1799,0.0866)	14.6246	0.3049	24.5931
	(0.9252,-0.1766,0.3358)	15.1033	0.3407	33.2576
	(-0.9774,-0.1819,0.1073)	20.4379	0.3544	20.8737
MS8 (n=798)	(-0.9354,0.1384,0.3255)	18.9917	0.2932	23.8760
	(-0.9985,-0.0160,-0.0531)	19.9218	0.3308	19.3526
	(0.9280,-0.1162,0.3541)	16.2867	0.3759	39.7512
MS9 (n=793)	(0.9577,0.0040,0.2877)	12.1347	0.3304	26.3245
	(-0.9788,0.1487,0.1409)	15.4375	0.3329	22.7304
	(-0.9520,-0.2417,0.1879)	14.5875	0.3367	30.8356

Table 4.5: The estimated values of the parameters of the vmf mixture model for the nine MS subjects (the first three columns). The angles (last column) between the mean directions are ordered as the angle between $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_2$, $\hat{\boldsymbol{\mu}}_2$ and $\hat{\boldsymbol{\mu}}_3$ then $\hat{\boldsymbol{\mu}}_3$ and $\hat{\boldsymbol{\mu}}_1$.

(the MS subjects have higher variance of FA). The test used is Mann-Whitney test.

The result in this chapter used the Euclidean method for segmenting the CC as it is faster than the segmentation methods using non-Euclidean methods (see Chapter 3). Using the segmentation results obtained by non-Euclidean methods can be explored in future work. Using the vmf to obtain similar results as Witelson subdivisions of the CC Witelson (1989) is also of interest.

Modeling diffusion directions of the CC using vmf distribution has additional benefits in that it can be used to automate a value for stopping criteria utilizing the max angle between the mean directions. This will be investigated in the next chapter.

Chapter 5

Fiber Tracking of the CC

Fiber tracking is a non-invasive technique to trace neural fiber pathways in the brain's white matter. Fiber tracking provides useful information about the brain anatomy and the connectivity in the brain which are useful in detecting various brain disorders. The noise in diffusion weighted imaging (DWI) and lesions in the brain affect deterministic fiber tracking by causing fibers to deviate and continue propagation out of the original pathways. In this chapter, we propose and evaluate an individual-based curvature threshold which is based on the largest angle between the mean directions obtained from modeling diffusion directions in the CC using von Mises-Fisher (vmf) distribution. Using the proposed curvature threshold, the fiber is terminated to prevent the deviation out of the correct pathway. Three common fiber tracking techniques (Fiber Assignment by Continuous Tracking (FACT), Bayesian and Wild Bootstrap) are used for constructing fiber tracts in healthy and MS subjects. The measures: fractional anisotropy (FA), mean diffusivity (MD) and length of fibers are used to compare fiber tracking quantitatively between the healthy and the MS subjects. Significant differences of length of fibers are obtained using the three fiber tracking techniques. In addition, there are insufficient evidences to suggest that any of the three measures obtained by using the three tracking methods is different. In this chapter, fiber tracking is introduced and deterministic and probabilistic tracking are discussed in Sections 5.1, 5.2 and 5.3. Stopping criteria is proposed and applications of fiber tracking are provided in Sections 5.4 and 5.5. The conclusion and discussion of the chapter are presented in Section 5.6.

5.1 Introduction and Related Work

Fiber tracking is able to detect many abnormalities in the brain. Mainly there are two types of fiber tracking algorithms, deterministic and probabilistic fiber tracking. Deterministic fiber tracking does not provide any information about the uncertainty of the tracked fibers due to errors in estimating directions. Probabilistic fiber tracking is more robust as the likelihoods of the fiber's pathways are considered by using a probability density function (pdf) to estimate the uncertainty of the fiber directions. However, one of the issues is that probabilistic fiber tracking that use the Monte Carlo method is affected by the distance from the starting point of tracking (seed point). The probability of fiber connectivity with the seed point is decreased as the distance increases (see Morris, Embleton, and Parker (2008)). Another issue is the large number of iterations needed in probabilistic fiber tracking to calculate the pdf. With high volumes of brain images, probabilistic fiber tracking for the whole brain images is very time consuming.

Stopping criteria of fiber tracking are needed to stop the propagation of fibers out of the white matter or out of a ROI in the brain, and to restrict the curvature of the fibers. The most common stopping criteria are anisotropy and curvature threshold. To restrict the anisotropy, a threshold value of any anisotropy measure can be used such as FA or relative anisotropy (RA). The curvature of fibers in most brain regions is low except some regions such as the Meyers loop portion of the optic pathway and the subcortical U-fibers, where a high curvature threshold such as 90 degrees is required (D. K. Jones, 2010). High curvature might allow a fiber to propagate back to a pre-visited track or might provide sharp pathways. The curvature is measured over a specified interval (called radius of curvature or curvature interval) and the fiber is terminated if the curvature exceeds the threshold over a specified curvature interval. P. J. Basser et al. (2000) suggest the curvature interval to be less than the length of two voxels, e.g. if the voxel resolution is $2.5 \times 2.5 \times 2.5$, then the suggested value for curvature interval is less than $5mm$. The default value of curvature interval in Camino toolkit is $5mm$.

Many researchers used a fixed curvature threshold for tracing fibers in a ROI or in the whole white matter of the brain. Examples of fixed curvature thresholds used are 80° (Behrens et al., 2003; Campbell, Siddiqi, Rymar, Sadikot, & Pike, 2005), 45° (Heemskerk, Sinha, Wilson, Ding, & Damon, 2010), 40° (Jiang, Van Zijl, Kim, Pearlson, & Mori, 2006), 35° (Marcella Laganà et al., 2012), 27° (Christidi, Karavasilis, Samiotis, Bisdas, & Papanikolaou, 2016). The curvature threshold used by default in Camino toolkit (Cook et al., 2006) and FMRIB software library (FSL) are 90° and 100° respectively. Using a fixed threshold for stopping criteria can lead to an overestimate or underestimate of tracking results Brecheisen et al. (2009). Marcella Laganà et al. (2012) proved that using an individual-based anisotropy threshold for tracing fibers of the CC in MS subjects is more reliable than using an atlas-based method. Noise in DWI data causes inaccurate estimation of diffusion tensors which can cause changes in the direction of principal eigenvectors. In addition, the direction of principal eigenvectors on lesions can be mistaken as the directional information is lost in these regions. These changes might lead fibers to deviate suddenly from the original pathways. Hence, a curvature threshold based on an individual's information is expected to provide a more accurate fiber tracking than a fixed curvature threshold, especially for deterministic tracking (as probabilistic tracking is less affected by noise).

Deterministic fiber tracking (e.g. Ozturk et al., 2010; Shu et al., 2011) and probabilistic fiber tracking (e.g. Pine, Jones, Lowe, Sakaie, & Phillips, 2009; Preziosa et al., 2011; Li, Jewells, et al., 2013; Lyksborg et al., 2014) have been used in MS and even combined (e.g. Klein et al., 2010; Hu et al., 2011). Shu et al. (2011) used Fiber Assignment by Continuous Tracking (FACT) algorithm to track fibers in group of MS subjects and the results are compared with healthy controls. They used graph theoretical methods to study the structural networks of the white matter in MS. Efficiency of the networks was found to be significantly reduced in MS subjects. Pine et al. (2009) showed that probabilistic fiber tracking is able to propagate through the lesions when using *spherical deconvolution* method for estimating fiber orientation.

Anatomical connectivity mapping (ACM) is a quantitative measure of the anatomical connectivity measured in each voxel in the brain (Embleton, Morris, Haroon, Ralph, & Parker, 2007). It combines information from probabilistic fiber tracking and anisotropy indices. Lyksborg et al. (2014) investigated the ACM and FA in Secondary Progressive and Relapsing-Remitting MS subjects. They demonstrated that ACM is more effective in distinguishing between Secondary Progressive MS and Relapsing-Remitting MS. This was due to reduction of ACM values in motor-related white matter in Secondary Progressive MS as compared to Relapsing-Remitting MS, while the decrease of FA values was not consistent. In addition, the correlation between ACM and the Expanded Disability Status Scale (EDSS) was stronger than the correlation between the FA and EDSS. Hu et al. (2011) found that both probabilistic and deterministic tracking produce less fibers in MS subjects. In addition, EDSS was correlated negatively with the fiber's density (as the disability increased, the density of the fibers decreased) for both probabilistic and deterministic tracking.

Much research has been conducted to explore fiber tracking of the CC in MS. Hasan et al. (2005) used FACT algorithm to study the FA of normal-appearing CC in relapsing-remitting MS. The FA was calculated through a seven segment of the CC (rostrum, genu, rostral midbody, anterior midbody, posterior midbody, isthmus and splenium). They found that while FA is significantly reduced in the anterior and posterior mid body subdivisions, the reduction of FA in the splenium was not significant. Fiber tracking can be interrupted by the lesions on the CC. This interruption is dependent on the size of the lesions in the CC (Lagana et al., 2009). Mesaros et al. (2009) studied the CC damage and its relationship to cognitive impairment in Benign MS using probabilistic fiber tracking. Kern, Sarcona, Montag, Giesser, and Sicotte (2011) used a probabilistic fiber tracking method to study the motor tracks in relapsing-remitting MS. They found a relationship between increased *radial diffusivity* in callosal motor fibers (the midbody of the CC) and hand weakness. Wahl et al. (2011) used the tensor deflection algorithm (Lazar et al., 2003) for tracking motor callosal fibers. They demonstrated the significant

reduction of the FA values in early stages of relapsing-remitting MS (i.e. before the development of a macroscopic lesion). In the CC, cerebellum and right hippocampus, ACM was correlated with Paced-Auditory-Serial-Addition-Test (PASAT) scores in Relapsing-remitting MS (Bozzali et al., 2013).

5.2 Deterministic White Matter Tractography

The deterministic tractography algorithms use diffusion tensor information to determine the fiber's orientation. There are two main techniques: streamlines, which assumes that the principal eigenvector of the diffusion tensor is parallel to the local fiber tract direction, and tensor deflection which use the entire diffusion tensor to construct the fiber pathways. Streamlines technique for fiber tracking was proposed by P. J. Basser (1998). Suppose the fiber pathway is represented by the space curve $\mathbf{r}(s)$, where s is the arc length. The main assumption is that the principal eigenvector $\mathbf{v}_1(\mathbf{r}(s))$ of the diffusion tensor at a voxel is equal to the local tangent vector of $\mathbf{r}(s)$ as follows:

$$\frac{d}{ds}\mathbf{r}(s) = \mathbf{v}_1(\mathbf{r}(s)) \quad (5.1)$$

The integration of Equation 5.1 is not analytical and a numerical method (e.g. Euler or Runge-Kutta) is needed to approximate $\mathbf{r}(s)$ (P. J. Basser et al., 2000). Although the Euler method is easy to implement, it is not accurate for second order derivatives or higher. In addition, the Euler method uses a fixed step size along the direction of the principal eigenvector in each voxel. The step size need to be significantly small to provide an accurate tracking. For example, P. J. Basser et al. (2000) indicate that any step size smaller than $0.05 \times l$, where l is the voxel dimension is sufficient. The Runge-Kutta method provides more accurate results than the Euler method, as it enables using a higher order correction scheme (P. J. Basser, 1998). In addition, the step size can be adaptive using the Runge-Kutta method (P. J. Basser et al., 2000). An interpolation method (e.g. linear interpolation or nearest-neighbour interpolation) is needed by both Euler and Runge-Kutta to provide continuous fiber pathways.

Another streamlines algorithm is FACT, which is proposed by Mori, Crain, Chacko, and Van Zijl (1999). The FACT method creates fiber tracts by following the direction of the principal eigenvectors continuously (i.e. at the boundary of a voxel the track should follow the principal eigenvector of the neighbour voxel). They showed that following the direction of the principal eigenvectors in a discrete way leads to the diversion of the fibers from the true fiber pathway. No interpolation is needed using the FACT algorithm as the step size is extended to the boundary of the voxel.

Comparing the FACT with the Runge-Kutta and the Euler algorithms, the Runge-Kutta and the Euler algorithms provide more smoother fiber tracking than the FACT algorithm (see Figure 5.1).

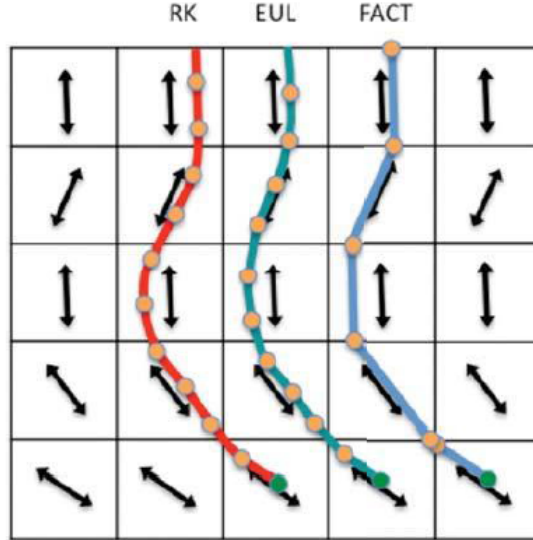


Figure 5.1: Fiber tracking using Runge-Kutta, Euler and FACT algorithm. Interpolation is needed for both Euler and Runge-Kutta methods. Small step size leads to smooth tracking (taken from (D. K. Jones, 2010)). Permission to reproduce this figure has been granted by Oxford University Press (Books).

The tensor deflection (TEND) is a deterministic tracking algorithm (Lazar et al., 2003). It uses the entire diffusion tensor to determine fiber direction. The

fiber track direction \mathbf{v}_{out} passing through a voxel is estimated using the diffusion tensor \mathbf{D} in the voxel as follows:

$$\mathbf{v}_{out} = \mathbf{D}\mathbf{v}_{in}, \quad (5.2)$$

where \mathbf{v}_{in} is the incoming vector i.e. the fiber direction from the previous step. The fiber directions at seed points are considered to be the same direction as the principal eigenvectors of the tensors at the seed points. The deflection of fiber direction towards the direction of the principal eigenvector depends on the degree of the anisotropy of the tensor (Lazar et al., 2003). The more anisotropy the tensor exhibits, the more deflected the fiber direction is towards the direction of the principal eigenvector. Lazar et al. (2003) showed that when the previous fiber direction is parallel or perpendicular to the principal eigenvector, the fiber direction is the same as the direction of the principal eigenvector (see Figure 5.2).

Another algorithm for tracking fibers is tensorline algorithm (Lazar et al., 2003; Weinstein, Kindlmann, & Lundberg, 1999) which is a general case of TEND. Suppose f and $g \in [0, 1]$ are user-defined weights, the direction of fiber track is

$$\mathbf{v}_{out} = f\mathbf{v}_1 + (1 - f)((1 - g)\mathbf{v}_{in} + g\mathbf{D}\mathbf{v}_{in}), \quad (5.3)$$

where \mathbf{v}_1 is the principal eigenvector. Equation 5.2 is a special case of Equation 5.3 when $f = 0$ and $g = 1$.

5.3 Probabilistic Tractography

Probabilistic tracking use a pdf to estimate the uncertainty of the fiber orientations due to systematic and random errors. There are two approaches for generating the pdf: parametric and nonparametric. Parametric approaches make assumptions on the possible noise on the model. Mainly there are two different parametric techniques for fiber tracking: data-based and model-based. In data based techniques, the probabilities are calculated from the DWI data using Bayesian approaches and

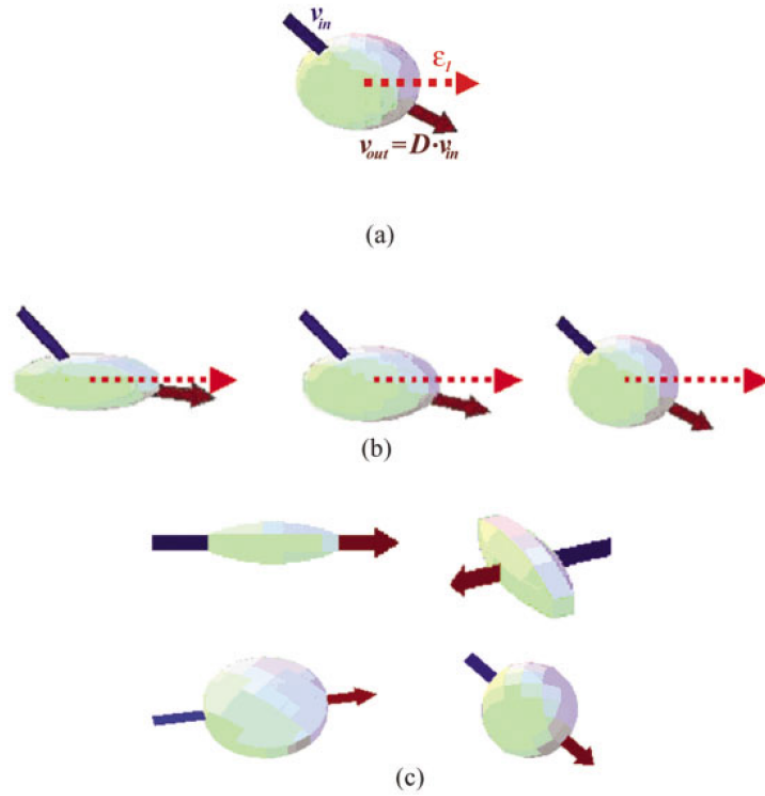


Figure 5.2: Tensor deflection algorithm for fiber tracking is shown in (a) where the dashed red line represents the principal eigenvector. The deflection of the fiber direction towards the principal eigenvector is related to the anisotropy of the tensor as shown in (b). In some cases, (c), the fiber direction is exactly same as the principal eigenvector (taken from Lazar et al. (2003)). Permission to reproduce this figure has been granted by John Wiley and Sons.

the posterior distribution is used as the pdf. In model-based techniques, a pdf (e.g. Gaussian, Watson, Bingham densities) is used to generate the probabilities. Non-parametric approaches (i.e. using bootstrap methods) make no assumptions on the underlying noise in the model.

5.3.1 Probability Density Function (pdf)

Behrens et al. (2003) proposed a Bayesian estimation method to quantify the uncertainty of the parameters on diffusion model. Samples are drawn from the joint posterior distribution using Markov chain Monte Carlo (MCMC). This method has then been applied to estimate the connectivity between two points in the human brain. Friman et al. (2006) proposed an alternative Bayesian method with simple and fast calculation for the posterior distribution by setting Dirac priors for nuisance parameters (parameters which are not of direct interest).

The bootstrap technique is a nonparametric method which does not make any assumption regarding the noise. The use of a bootstrap method (Efron, 1992) for fiber tracking is proposed by many authors (e.g. Lazar & Alexander, 2005; D. K. Jones & Pierpaoli, 2005; D. K. Jones, Travis, Eden, Pierpaoli, & Basser, 2005). They used a deterministic white matter tractography (e.g. second order RungeKutta streamlines) with bootstrap sampling. Multiple measurements are needed in each diffusion direction for bootstrap analysis. Whitcher et al. (2008) proposed using the Wild Bootstrap (Flachaire, 2005; Liu et al., 1988) for fiber tracking where one measurements is needed per diffusion direction. They showed that using this method provides the same level of accuracy as provided by the regular bootstrap methods.

The pdf can be a Gaussian distribution or a directional distribution function (e.g. Watson, Bingham). Parker and Alexander (2003) used Gaussian distribution, with 0 mean and signal to noise ratio (SNR) similar to the SNR of the real data, for noise simulation.

Fiber tracking can be local where the pdf is defined using voxel information and the neighbouring information is ignored. Global tracking uses the information from the voxel's neighbours to track fibers (e.g. Kreher, Mader, & Kiselev, 2008; Jbabdi, Woolrich, Andersson, & Behrens, 2007).

5.3.2 Probabilistic Local Tracking Methods

Once a pdf for characterizing the uncertainty of fiber pathways is determined at each voxel, the possible track pathways can be identified using a streamline approach with Monte Carlo simulations (Parker, Haroon, & Wheeler-Kingshott, 2003). A random sample is drawn from the pdf, at each voxel, and repeated many times to construct the possible tracks pathways. The probability that a fiber passes through a voxel at position r can be defined approximately as the number of times that a fiber passes through this voxel $M(\mathbf{r}, N)$ divided by the total number of the Monte Carlo iterations N (Parker et al., 2003) as follows:

$$P(\mathbf{r}) = \frac{M(\mathbf{r}, N)}{N} \quad (5.4)$$

Other alternative methods of Monte Carlo streamlines for fiber tracking are random walks using virtual particles (e.g. Koch, Norris, and Hund-Georgiadis (2002)) and front propagation methods (e.g. Parker, Wheeler-Kingshott, and Barker (2002)). See D. K. Jones (2010) for more details.

5.4 Stopping Criteria

Generally the termination criteria of fiber tracking method are based upon curvature and anisotropy thresholds. A binary brain mask can also be used to stop tracing fibers as an alternative to anisotropy threshold. In addition, the tracking might stop if the length of the fiber exceeds a specific value (e.g. for Camino toolkit the tracking stops if the fiber length exceed $1000mm$).

A sudden change of the propagation pathways of a fiber could be related to the noise in DWI data, which causes directional change in the true direction of

principal eigenvectors (P. J. Basser et al., 2000) and also could be related to the destruction of the directional information due to lesions (in MS subjects). In Chapter 4, the diffusion of the CC is modeled using a mixture of vmf distributions. This modeling of the diffusion directions in the CC provides three different clusters of diffusion directions, and hence the diffusion directions in each cluster are similar (i.e. small angles between the principal eigenvectors inside each cluster). However, the diffusion directions in between the clusters are significantly different. Hence, the angles between the mean directions can provide useful information about the curvature of diffusion direction in the CC. We used the three angles between the mean directions to propose the measure sc for the curvature threshold as stopping criteria for tracing fibers in the CC as follows:

$$sc = \max(\theta_1, \theta_2, \theta_3), \quad (5.5)$$

where θ_1 , θ_2 and θ_3 are the angles between the three mean directions. This measure is evaluated using a group of healthy and MS subjects (see Section 5.5).

5.5 Application

The data used for fiber tracking in this study are nine secondary progressive MS and four healthy subjects (which is the same data used in Chapter 4). For details about the data and the imaging protocol, see Section 4.4.2. Diffusion tensors are fitted linearly (P. Basser, Mattiello, & LeBihan, 1994). The tracking is performed using Camino toolkit (Cook et al., 2006). Data is then analyzed using Matlab and Minitab. For all the subjects, the seed points are chosen from the sagittal plane on the middle of the CC with 38 voxels.

Subject	sc
H1	37.1
H2	40.3
H3	32.7
H4	41.7
MS1	34.0
MS2	32.4
MS3	29.1
MS4	26.7
MS5	26.4
MS6	34.4
MS7	33.3
MS8	39.8
MS9	30.8

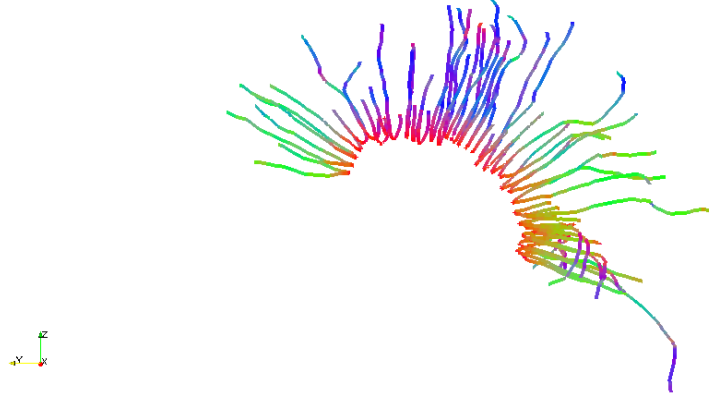
Table 5.1: The curvature threshold sc for four healthy and nine MS subjects.

5.5.1 Evaluation of Stopping Criteria

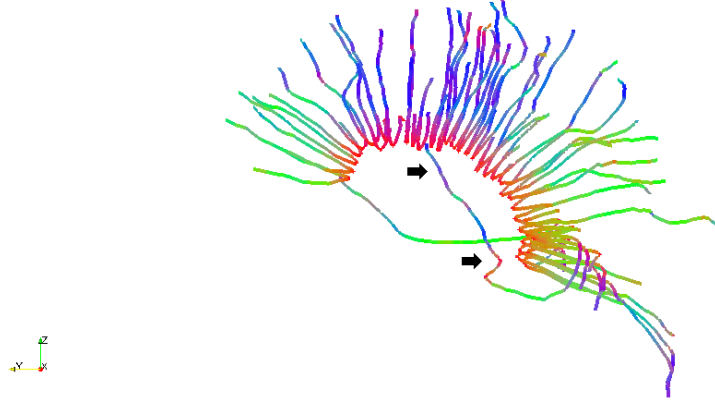
To evaluate the proposed stopping criteria (Equation 5.5), we perform fiber tracking of the CC using FACT tracking method. To enable the evaluation of sc without interaction of other stopping criteria, the only stopping criteria used is the curvature threshold (the same results are also obtained when adding an anisotropy threshold i.e. $FA = 0.2$). The values of sc for all the subjects are shown in Table 5.1. Then we perform fiber tracking using a fixed curvature threshold 45° , as the values of sc are all less than 45° . Comparing the results of fiber tracking using sc and 45° , we find that both curvature thresholds provide fibers of the CC without sharp deviation from the original pathways in the four healthy subjects (e.g. see Figures 5.3a and 5.4a), with more propagation of some fibers out of the CC boundary when using 45° in three of the healthy subjects (e.g. Figures 5.3b and 5.4b). For the rest of the healthy subjects see Appendix A.4.

For MS subjects, we find that using sc does not allow fibers to deviate sharply from the original pathway which results in termination of the fiber where there is abnormality (e.g. see Figures 5.5a and 5.6a). The deviation can be seen by sudden change of the fiber directions e.g. change of the color from red to green in the middle of the CC (e.g. see the black stars in Figures 5.5b and 5.6b). For the rest of the MS

subjects see Appendix A.4.



(a) Fiber tracking of the CC using the curvature threshold $sc = 37.1^\circ$.



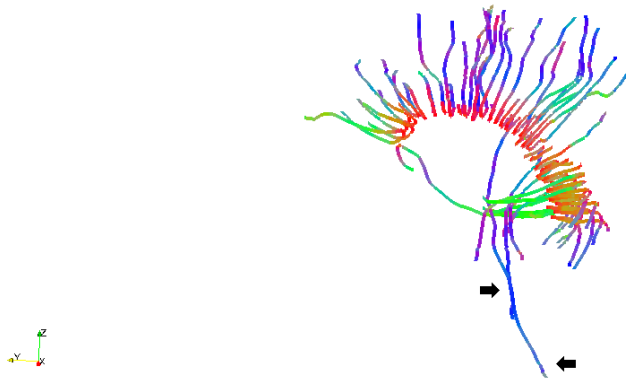
(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure 5.3: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the healthy subject H1. One fiber continues propagation out of the CC (indicated with the black arrows).

The voxel resolution for the brain images is 2.3 mm^3 (see Chapter 4), and hence the suggested curvature interval is less than $2 \times 2.3 = 4.6 \text{ mm}$ (P. J. Basser et al., 2000). The curvature interval used for all fiber tracking results is 1 mm for both curvature thresholds sc and 45° . That is because, for both curvature thresholds, a higher curvature interval, e.g. 2 mm , causes deviation of some fibers out of the original tracts. For example, fiber tracking of the CC of MS1 and MS6 using the curvature threshold sc and 45° (and the two curvature intervals 1 mm and 2 mm)

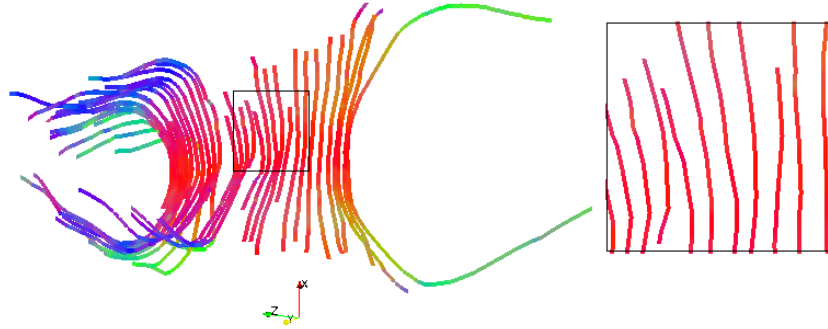


(a) Fiber tracking of the CC using the curvature threshold $sc = 32.7^\circ$.

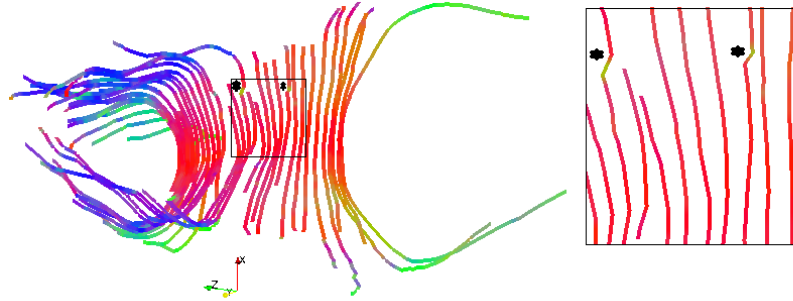


(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure 5.4: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the healthy subject H3. Using 45° results in continued propagation of some fibers out of the CC (indicated with the black arrows).

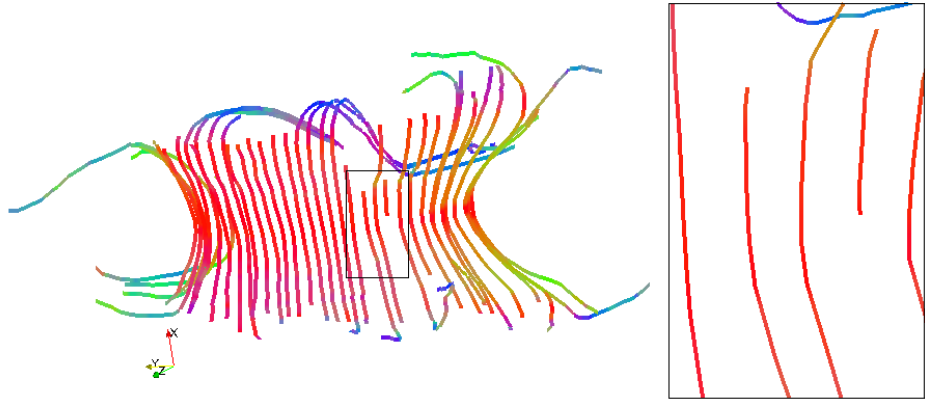


(a) Fiber tracking of the CC using the curvature threshold $sc = 34^\circ$.

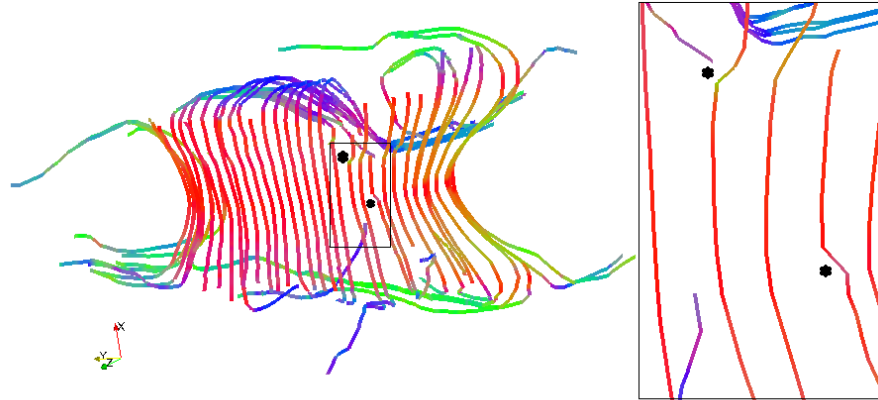


(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure 5.5: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the MS subject MS1. Using 45° results in deviation of some fibers from the original pathways (indicated by the black stars and zoomed in to the shape on the right).



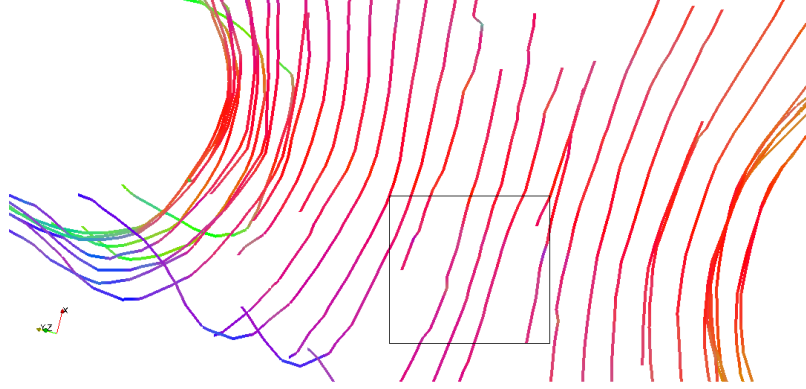
(a) Fiber tracking of the CC using the curvature threshold $sc = 34.4^\circ$.



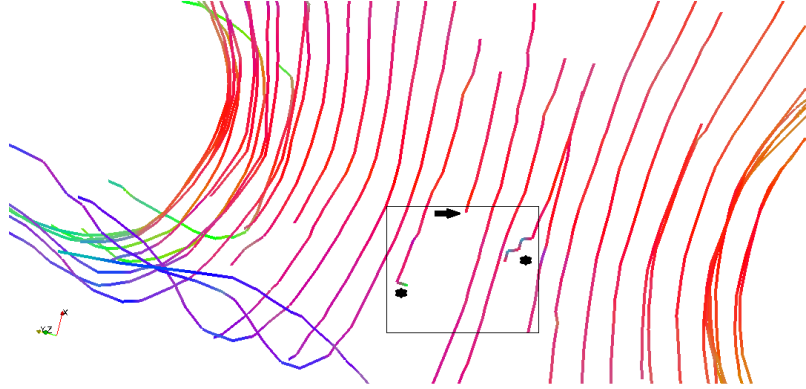
(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure 5.6: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the MS subject MS6. Using 45° results in deviation of some fibers from the original pathways (indicated by the black stars and zoomed in to the shape on the right).

are shown in Figures 5.7, 5.8, 5.9 and 5.10.



(a) Fiber tracking of the CC using the curvature interval 1 *mm*.



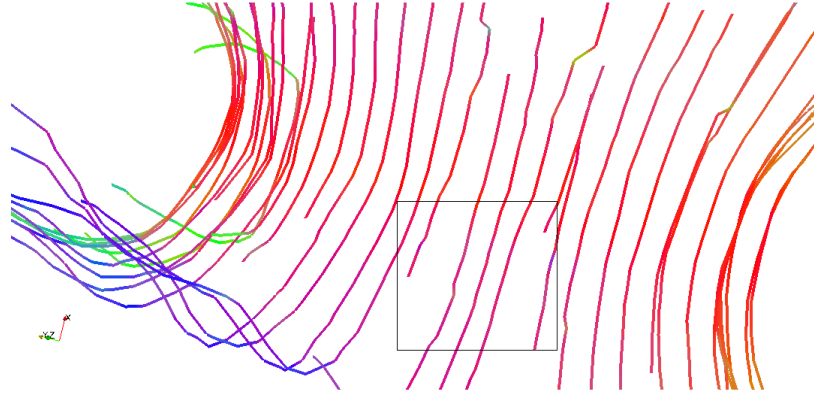
(b) Fiber tracking of the CC using the curvature interval 2 *mm*.

Figure 5.7: Comparing fiber tracking of the CC using two curvature intervals and the same curvature threshold sc in the MS subject MS1. Using curvature interval 2 *mm* results in deviation of some fibers from the original pathways (indicated by the black stars) and premature termination of one fiber (indicated by the black arrow).

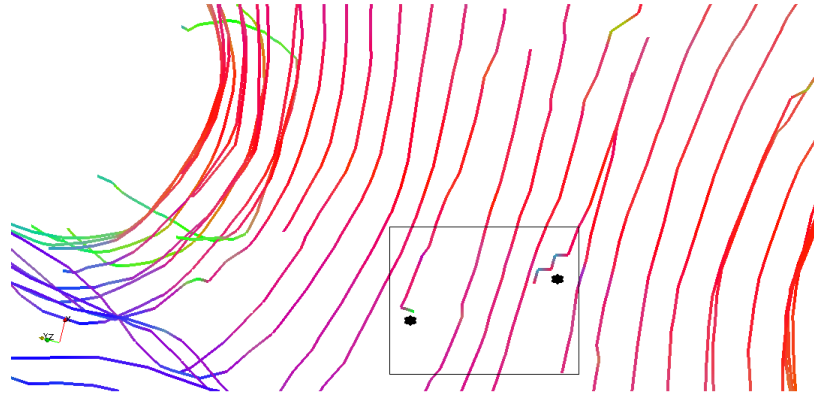
The proposed curvature threshold is used as stopping criteria for the application of fiber tracking of the CC in the next section.

5.5.2 Deterministic and Probabilistic Tracking

In this section, we compare deterministic and probabilistic fiber tracking quantitatively using three of the common used fiber tracking methods: Bayesian (Friman et al., 2006) and Wild Bootstrap (Whitcher et al., 2008) (probabilistic tracking meth-

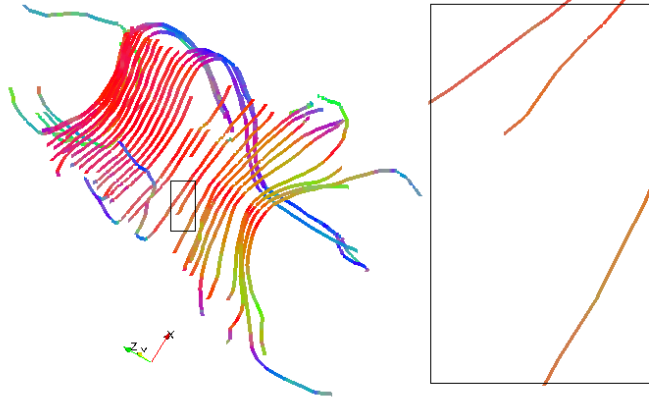


(a) Fiber tracking of the CC using the curvature interval 1 *mm*.

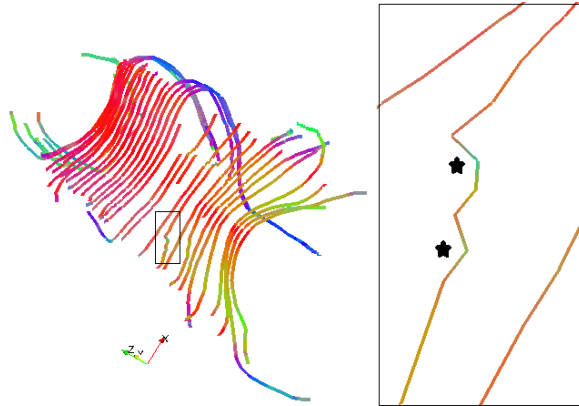


(b) Fiber tracking of the CC using the curvature interval 2 *mm*.

Figure 5.8: Comparing fiber tracking of the CC using two curvature intervals and the same curvature threshold 45° in the MS subject MS1. Using curvature interval 2 *mm* results in deviation of some fibers from the original pathways (indicated by the black stars).

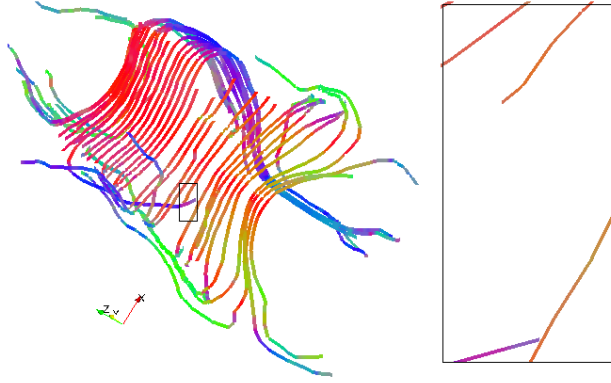


(a) Fiber tracking of the CC using the curvature interval 1 *mm*.

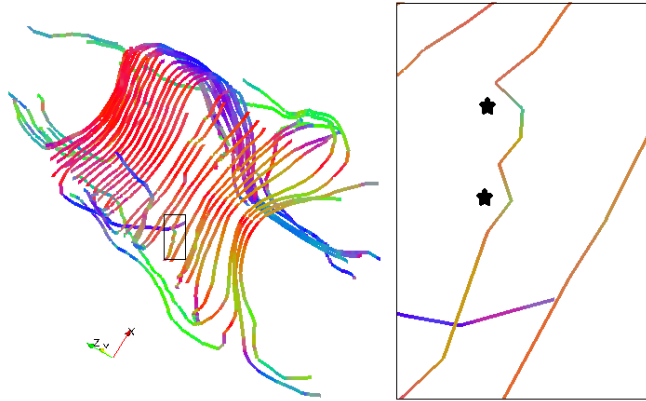


(b) Fiber tracking of the CC using the curvature interval 2 *mm*.

Figure 5.9: Comparing fiber tracking of the CC using two curvature intervals and the same curvature threshold sc in the MS subject MS6. Using curvature interval 2 *mm* results in deviation of a fiber from the original pathway (indicated by the black stars).



(a) Fiber tracking of the CC using the curvature interval 1 *mm*.



(b) Fiber tracking of the CC using the curvature interval 2 *mm*.

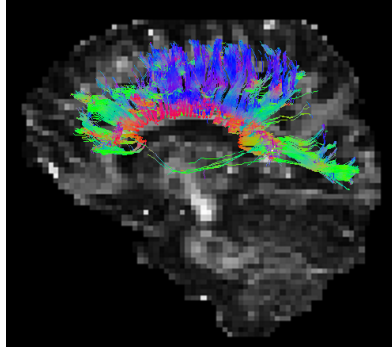
Figure 5.10: Comparing fiber tracking of the CC using two curvature intervals and the same curvature threshold 45° in the MS subject MS6. Using curvature interval 2 *mm* results in deviation of a fiber from the original pathway (indicated by the black stars).

ods) and FACT (deterministic tracking method). For Bayesian method, the prior information used are that S_0 (the Magnetic Resonance signal without the diffusion weighting) and variance are non-negative, as there is no detailed prior information available. The results of tracking might change depending on the prior information used. Fiber tracking of the healthy and the MS subjects are also compared. Probabilistic fiber tracking methods are performed using Monte Carlo simulations with 1000 iterations. The Fractional anisotropy (FA) threshold is chosen as 0.13 which is smaller than the minimum value of FA in the CC for both MS and healthy subjects. The results of fiber tracking for one MS and one healthy subject are shown in Figure 5.11. It is clear that the fibers are longer for the healthy than the MS subject (see the yellow arrows).

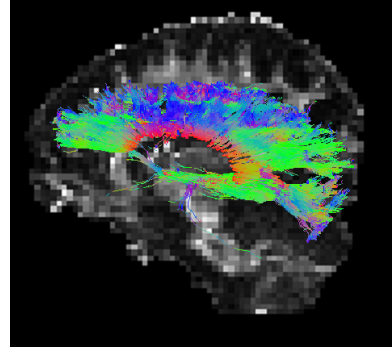
Results

To evaluate the results of fiber tracking of the CC quantitatively for the healthy and MS subjects using deterministic and probabilistic tracking, the average of length, FA and MD of each streamline in the CC are calculated. For testing the difference between the tracking methods, we used Anova to test the data that satisfies Anova conditions (normality of residuals and equal variances) otherwise Kruskal-Wallis test is used for testing non-parametric data. The results of the test show insufficient evidence to suggest that the FA, MD and length of the fibers obtained by the three tracking methods are different for both healthy (Table 5.2) and MS (Table 5.3) subjects at $\alpha = 0.05$. Generally, probabilistic tracking methods provide longer fibers than FACT, for most of the subjects (see the last column in Tables 5.2 and 5.3).

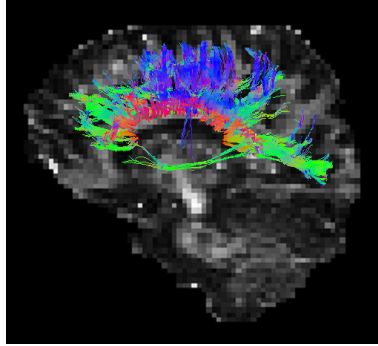
To test the difference between the healthy and MS subjects, we calculate the average of each of the measures (FA, MD and length) of each subject and used Mann-Whitney test (Table 5.4) as the sample size is small (4 values for healthy and 9 values for MS). The test yields sufficient evidences to suggest that length of the



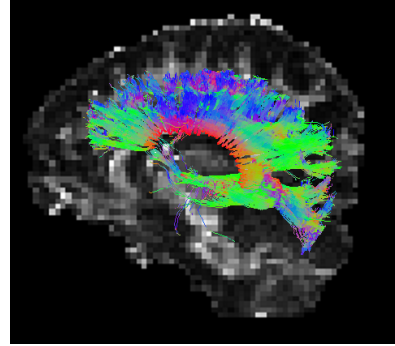
(a) Bayesian tracking of MS subject



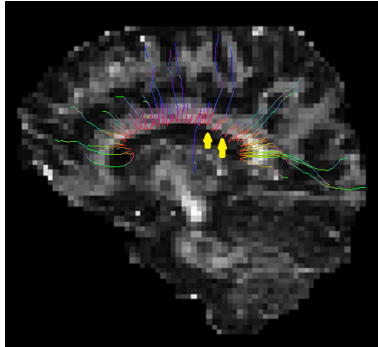
(b) Bayesian tracking of healthy subject



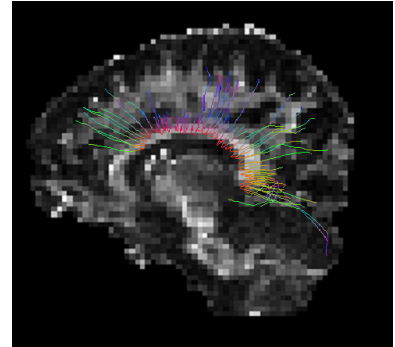
(c) Wild Bootstrap tracking of MS subject



(d) Wild Bootstrap tracking of healthy subject



(e) FACT tracking of MS subject



(f) FACT tracking of healthy subject

Figure 5.11: Fiber tracking of the CC in MS subject (right column) and healthy subject (left column) using Bayesian, Wild Bootstrap and FACT tracking methods. Lesions are much clearer in FACT figure (indicated with yellow arrows).

fibers in the CC of healthy and MS subjects are significantly different (at $\alpha = 0.05$) using the three tracking methods. In addition, the test indicates insufficient evidences to suggest that FA and MD are different in the healthy and the MS subjects at $\alpha = 0.05$. The healthy subjects are old (mean age \pm sd = 62.80 ± 9.50) and the MS subjects are relatively young (mean age \pm sd = 43.60 ± 10.07). FA and MD are both affected by age and MS and hence the differences between the healthy and the MS subjects are not significant.

Subjects	Tracking	avg FA	avg MD $\times 10^9$	avg length
H1	Bayesian	0.6479	0.7567	61.35
	Bootstrap	0.6367	0.7568	65.29
	FACT	0.6324	0.7489	66.00
	p-value	0.983	0.948	0.867
H2	Bayesian	0.4897	0.8443	57.29
	Bootstrap	0.4916	0.8425	55.15
	FACT	0.5001	0.8302	53.00
	p-value	0.386	0.864	0.135
H3	Bayesian	0.5970	0.9439	50.22
	Bootstrap	0.5862	0.9011	52.44
	FACT	0.5828	0.8878	49.50
	p-value	0.868	0.825	0.628
H4	Bayesian	0.6277	0.7844	66.43
	Bootstrap	0.6260	0.7871	65.85
	FACT	0.6271	0.7570	62.00
	p-value	0.898	0.412	0.307

Table 5.2: Average of FA, MD (mm^2/sec) and length (mm) of the fibers in the CC in four healthy subjects. The p-values indicate that there are insufficient evidences to suggest differences between the three tracking methods in length, FA and MD along the CC fibers. The test used is Anova (when normality of residuals and equal variances conditions are satisfied) otherwise Kruskal-Wallis test is used.

5.6 Conclusions and Discussions

We proposed a measure of curvature threshold sc , as stopping criteria for fiber tracking of the CC based on the angles between the mean directions in the CC. For the healthy subjects, using sc reduces the number of the fibers that propagate out of

Subjects	Tracking	avg FA	avg MD $\times 10^9$	avg length
MS1	Bayesian	0.51495	0.8694	48.25
	Bootstrap	0.51702	0.8658	49.70
	FACT	0.51961	0.8640	46.50
	p-value	0.946	0.854	0.864
MS2	Bayesian	0.6604	0.8142	54.63
	Bootstrap	0.6559	0.8106	54.64
	FACT	0.6411	0.8181	50.00
	p-value	0.850	0.973	0.900
MS3	Bayesian	0.6253	0.8486	45.46
	Bootstrap	0.6199	0.8522	47.86
	FACT	0.6264	0.8361	44.00
	p-value	0.902	0.936	0.379
MS4	Bayesian	0.5603	0.9505	52.77
	Bootstrap	0.5692	0.9484	53.39
	FACT	0.5554	0.9401	56.50
	p-value	0.736	0.852	0.331
MS5	Bayesian	0.6482	0.8578	48.80
	Bootstrap	0.6515	0.8580	47.82
	FACT	0.6526	0.8471	45.50
	p-value	0.963	0.993	0.259
MS6	Bayesian	0.6632	0.8246	47.33
	Bootstrap	0.6612	0.8304	48.49
	FACT	0.6766	0.8350	46.00
	p-value	0.794	1.000	0.276
MS7	Bayesian	0.5989	0.9279	51.89
	Bootstrap	0.6090	0.9317	52.07
	FACT	0.5990	0.9096	50.50
	p-value	0.944	0.971	0.439
MS8	Bayesian	0.6210	0.8010	50.04
	Bootstrap	0.6232	0.8034	51.19
	FACT	0.6248	0.8042	45.00
	p-value	0.960	0.830	0.490
MS9	Bayesian	0.6159	0.8779	46.48
	Bootstrap	0.6196	0.8841	48.26
	FACT	0.6095	0.8795	45.50
	p-value	0.989	0.971	0.953

Table 5.3: Average of FA, MD (mm^2/sec) and length (mm) of the fibers in the CC in nine MS subjects. All the p-values indicate that there are insufficient evidences to suggest differences between the three tracking methods in length, FA and MD along the CC fibers. The test used is Anova (when normality of residuals and equal variances conditions are satisfied) otherwise Kruskal-Wallis test is used.

Tracking	Subjects	avg FA	avg MD $\times 10^9$	avg length
Bayesian	Healthy	0.6123	0.8144	59.32
	MS	0.6210	0.8578	48.80
	CI	(-0.1313,0.0674)	(-0.1212,0.0861)	(1.42,17.63)
	P-value	0.4875	0.3159	0.0253
Bootstrap	Healthy	0.6061	0.8148	60.22
	MS	0.6232	0.8580	49.70
	CI	(-0.1316,0.0568)	(-0.1273,0.0431)	(1.25,17.43)
	P-value	0.5892	0.2472	0.0168
FACT	Healthy	0.6049	0.7936	57.50
	MS	0.6248	0.8471	46.00
	CI	(-0.1247,0.0688)	(-0.1306,0.0407)	(2.50,19.99)
	P-value	0.5892	0.1897	0.0372

Table 5.4: Average of FA, MD (mm^2/sec) and length (mm) of the fibers in the CC in the healthy and the MS subjects. The p-values and the confidence intervals (CI) for the differences indicate that the length of the CC fibers for the MS subjects are significantly different from the length of the CC fibers for the healthy subjects. The test used is Mann-Whitney test.

the boundary of the CC. For the MS subjects, using *sc* stop fibers from deviation into incorrect pathways. Hence, for MS and healthy subjects, the individual-based curvature threshold outperform the fixed curvature threshold (i.e. 45°) for fiber tracking in the CC. We showed that, for both curvature thresholds, fiber tracking using curvature interval as 1 mm outperforms fiber tracking using 2 mm . The evaluation of the curvature threshold is performed using FACT algorithm. Although the probabilistic tracking is expected to be more robust generally than the deterministic tracking and less affected by noise, more investigation will be considered in future work to evaluate the proposed curvature threshold using probabilistic tracking. Using an accurate segmentation method of the CC (e.g. the log Euclidean method in Chapter 3) can provide an accurate result of mean directions in the CC (Chapter 4) and hence an accurate curvature threshold of fiber tracking in the CC.

Furthermore, we compared the fiber tracking results of the CC quantitatively using two probabilistic (Bayesian and Wild Bootstrap) and one deterministic (FACT) tracking methods. The measured used for comparisons are FA, MD and length of the fibers. The results indicate that there are insufficient evidences to sug-

gest differences in the measures (FA, MD and length of the fibers) obtained using the three fiber tracking methods at $\alpha = 0.05$. However, probabilistic fiber tracking provides all the possible fiber pathways and hence the uncertainty of fiber tracts can be assessed using probabilistic fiber tracking.

When comparing the results of fiber tracking in the healthy and the MS subjects, the length of the fibers are significantly shorter in the MS subjects as compared with the healthy subjects (at $\alpha = 0.05$). The FA and MD values in the healthy subjects are not significantly different as compared with the MS subjects. The significant reduction of FA and increase in MD in MS subjects when compare with age-matched healthy subjects are well known in the literature (e.g. Tievsky et al., 1999; Filippi et al., 2001; Coombs et al., 2004). The healthy and MS subjects used in this work are not age-matched as the healthy subjects are relatively older than the MS subjects. In healthy subjects, the FA decreases and MD increases with aging (Burzynska et al., 2010; Sullivan, Rohlfing, & Pfefferbaum, 2010; Zahr, Rohlfing, Pfefferbaum, & Sullivan, 2009; Sullivan, Adalsteinsson, & Pfefferbaum, 2005). This confirms that FA and MD are both affected by the age and abnormalities and hence they can be used for distinguishing between healthy and MS subjects only if the groups consist of aged-matched participants. However, the length of fibers is a robust measure to distinguish between MS and healthy subjects even in a group of non age-matched participants.

Chapter 6

Conclusions and Future Work

The aim of the thesis was to develop statistical methodologies for accurate grouping and modeling data using Diffusion Tensor Imaging (DTI), taking into account the complex nature of the data, to improve disease diagnosis and surgical planning. In fact, the study focused on the Corpus Callosum (CC) as it is a very important region in the brain which integrates the information (i.e. sensor, motor and cognitive) between the two hemispheres of the brain and also changes to the size or the shape of the CC is an indicator of a brain disorder.

In this chapter, a brief summary of the research findings and conclusions of the thesis is provided in Section 6.1. Then discussions and directions of future works are presented in Section 6.2.

6.1 Summary and Conclusions

In this thesis, a general background of DTI is provided in Chapter 2. The Magnetic Resonance Imaging (MRI) and Diffusion Weighted Imaging (DWI) are first reviewed. Then DTI is discussed.

In Chapter 3, we proposed an accurate automated method for segmentation of the CC using Hartigan's method and DTI. Hartigan's method was generalized for the use with f -mean metrics to cluster diffusion tensors (and covariance matrices

generally which can be used, for example, in longitudinal data analysis). The generalized Hartigan’s method was then applied to segment the 2D and 3D CC of a healthy brain. For 2D and 3D segmentation of the CC, the log Euclidean and Riemannian methods provided more accurate segmentation than the other methods (Euclidean, Cholesky, root Euclidean and Procrustes). When compared with the manual segmentation, the log Euclidean and Riemannian methods provided the most accurate 2D segmentation. For 2D segmentation, they yielded the least variations in the size and shape of the tensors in the CC. For 3D segmentation, the log Euclidean and Riemannian methods were the only methods that segment all of the splenium (posterior end) as part of the CC. Hence, the log Euclidean and Riemannian methods outperformed the other methods (i.e. Euclidean, Cholesky, root Euclidean and Procrustes) in the segmentation of the CC. The log Euclidean method is more efficient, as the computation of the log Euclidean mean is much faster than the Riemannian mean (Arsigny et al., 2007). The computation of the Euclidean distance is faster than the computation of the non-Euclidean distances. Furthermore, the Riemannian and Procrustes means use iterative numerical methods for the computations and hence are computationally more expensive than the other means (Euclidean, Cholesky, root Euclidean and log Euclidean). To explore another clustering method, hierarchical clustering was also applied to segment the 2D CC. Compared with Hartigan’s method, hierarchical clustering (single link) yielded holes on the segmented 2D CC whereas no holes were produced when using Hartigan’s method. Because no holes are expected in the CC of a healthy brain, Hartigan’s method was preferred.

The von Mises-Fisher distribution (vmf) is a probability distribution for modeling directional data on the unit hypersphere. Diffusion patterns in the brain are affected by abnormalities (e.g. lesions) and hence analysing diffusion directions is useful in diagnosis of such abnormalities. In Chapter 4, we modeled the diffusion directions of the CC as a mixture of vmf distributions for Multiple sclerosis (MS) and healthy subjects. The segmentation of the CC using the Euclidean method (Chapter 3) was applied to segment the CC in a group of healthy and MS subjects due to the

faster computation of the Euclidean method as compared with the non-Euclidean methods. Bayesian information criterion (BIC) indicated that there were at least three different mean diffusion directions in the CC. Hypothesis tests confirmed that the three mean directions are significantly different from each other. Diffusion concentration around the mean directions were higher and the sum of angles between the mean directions were smaller on the normal-appearing CC of the MS subjects as compared to the healthy subjects. The variance of FA in MS subjects was significantly higher as compared with the healthy subjects. Curvature threshold is one of the stopping criteria for fiber tracking. The maximum angle between the three mean directions was proposed as an individual-based curvature threshold to provide stopping criteria for fiber tracking in the CC.

Fiber tracking is a non-invasive technique to understand the brain functions and the connectivity in the brain which is useful to diagnose various brain abnormalities. In Chapter 5, the proposed individual-based curvature threshold (Chapter 4) was evaluated using the healthy and MS subjects (the same data that was used in Chapter 4). Using the proposed curvature threshold, the number of fibers that propagated out of the CC of the healthy subjects was reduced, and the deviation out of the original fiber pathways in the CC of the MS subjects was prevented. Therefore, the individual-based curvature threshold was more preferable than a fixed curvature threshold for tracing fibers in the CC. In addition, three measures (average of: FA, MD, and length of fibers) obtained from fiber tracking of the CC were compared between three fiber tracking methods (FACT, Bayesian and Wild Bootstrap) and between the healthy and MS subjects at significance level $\alpha = 0.05$. There were insufficient evidences to suggest differences in the measures between the three tracking methods. There were significant differences in length of fibers between the healthy and the MS subjects (shorter length of the fibers in MS subjects). However, insignificant differences between the healthy and the MS subjects in FA and MD were obtained. These are due to the differences in age between the MS subjects (mean age \pm sd = 43.60 ± 10.07) and the healthy subjects (mean age \pm sd = 62.80 ± 9.50). FA

decreases and MD increases by both aging and MS. This demonstrates that FA and MD are both sensitive to the age and hence care should be taken when comparing between groups of a non aged matched participants using FA and MD. The length of fibers however is a robust measure to distinguish MS and healthy even using a non aged matched participants.

The relationships between the chapters are summarized as follows:

- **Chapter 3:** An accurate grouping method for DTI was developed taking into account the complex nature of the data. This method was then used to develop an accurate 2D and 3D segmentation of the CC.
- **Chapter 4:** The segmentation method in **Chapter 3** was used to segment the CC in a group of healthy and MS subjects. Then the diffusion directions of the CC were modeled using vmf distribution to obtain and analyze information which were used to propose individual-based curvature threshold.
- **Chapter 5:** Improved fiber tracking of the CC was obtained using the proposed curvature threshold and the same data in **Chapter 4**.

Overall, statistical methodologies for accurate grouping and modeling data using DTI were developed, taking into account the complex nature of the space of DTI.

6.2 Discussions and Future Work

In this section, some discussions regarding evaluation of the segmentation of the CC and diffusion directions are provided. In addition, some of the additional interesting areas of research to explore in the future are recommended.

6.2.1 Evaluation of the Segmentation of the CC

Using Hartigan's method (Chapter 3), the log Euclidean and Riemannian methods provided more accurate segmentation of the CC as compared with the other methods. This confirms and extends the results obtained by Lenglet et al. (2006) which

indicated that the Riemannian distance provides improved segmentation of the CC, especially in the regions of the splenium, as compared with the Euclidean distance. The data used was one healthy brain and the manual segmentation was provided by an expert. More data and multiple experts will be considered in future works to confirm the evaluation results.

6.2.2 Classification of Brain Tissue

Classification of brain tissues is assigning a tissue class to each voxel in the image (e.g. white matter, gray matter, cerebrospinal fluid, muscle, fat and air). Both classification and segmentation are partitioning of the image into homogeneous regions. Unlike the segmentation, the classes (regions) in the classification can be disconnected (Palm Dzung, Chenyang, & Prince Jerry, 1998). Palm Dzung et al. stated that the number of classes in an image are often based on prior information. The previous methods of classification of the brain are mostly based on MRI (e.g. Choi, Haynor, & Kim, 1991; Lachmann & Barillot, 1992; Bullmore et al., 1995; Sanjay-Gopal & Hebert, 1998; Van Leemput, Maes, Vandermeulen, & Suetens, 1999; Suckling, Sigmundsson, Greenwood, & Bullmore, 1999; Shattuck, Sandor-Leahy, Schaper, Rottenberg, & Leahy, 2001; Cocosco, Zijdenbos, & Evans, 2003; Suri, 2004; Awate, Tasdizen, Foster, & Whitaker, 2006; Hartlep, Pedain, Brady, & Raghavan, 2006). There has been little research on tissue classification using DTI. A classification and quantification method based on DTI has been proposed by Ding, Gore, and Anderson (2003) but it is limited to fiber pathways in the brain. A classification method using DTI can be explored in the future.

6.2.3 Clustering the Brain Lobes

The cerebral cortex of the brain is classified into four lobes: frontal lobe, parietal lobe, temporal lobe and occipital lobe. Using the spectral clustering (eigenfunctions of *Laplace-Beltrami Operator*) to classify the cerebral cortex into the brain lobes has been proposed by Lefevre, Auzias, and Germanaud (2014). Studying the clustering of the cerebral cortex using DTI (i.e. apply the segmentation method in Chapter 3

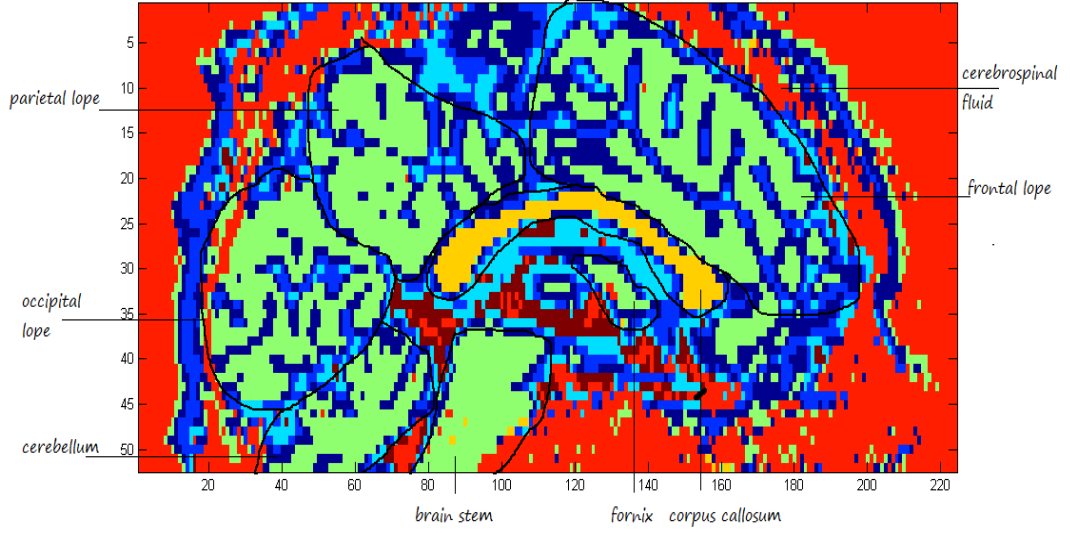


Figure 6.1: Segmentation of the CC of a whole slice of the brain using Euclidean method. The temporal lobe is not identified because the figure shows a midsagittal slice of the brain.

as in Figure 6.1) can be explored as another potential application to the developed clustering method.

6.2.4 Diffusion Directions

In Chapter 4, the Euclidean method was used for segmentation of the CC as it is faster than the segmentation using the non-Euclidean methods but it is less accurate than the log-Euclidean and Riemannian methods. The log-Euclidean or Riemannian methods can be applied to provide more accurate segmentation of the CC and hence more accurate parameters of modeling diffusion directions.

For MS subjects, the diffusion in the lesion regions are distributed randomly without restriction, and hence directional information is lost in those regions. The modeling results were obtained using the normal-appearing CC and lesions are not included to prevent the expected bias in the modeling results, if a misleading principal eigenvector is used. Future work can address how the modeling results can be affected when the regions with MS lesions in the CC are included in the analysis.

Furthermore, the age of the MS subjects did not match the age of the healthy subjects. The mean age of the MS subjects \pm sd is 43.60 ± 10.07 and the mean age of the healthy subjects \pm sd is 62.80 ± 9.50 . The results of FA showed no significant difference between the MS and the healthy subjects. This might be due to the age differences because the FA decreases with the age and hence both the MS and healthy subjects have similarly decreased FA. A larger data set with age matched participants can be considered in future work.

Appendix A

Appendices

A.1 Euclidean and non-Euclidean Distances

The Euclidean and non-Euclidean distances, between the covariance matrices \mathbf{A}_1 and \mathbf{A}_2 , have been summarized in Dryden et al. (2009) as follows:

Clustering method	distance
Euclidean	$\ \mathbf{A}_1 - \mathbf{A}_2\ $
Log Euclidean	$\ \log(\mathbf{A}_1) - \log(\mathbf{A}_2)\ $
Riemannian	$\left\ \log(\mathbf{A}_1^{-1/2} \mathbf{A}_2 \mathbf{A}_1^{-1/2}) \right\ $
Cholesky	$\ \text{chol}(\mathbf{A}_1) - \text{chol}(\mathbf{A}_2)\ $
Root Euclidean	$\left\ \mathbf{A}_1^{1/2} - \mathbf{A}_2^{1/2} \right\ $
Procrustes	$\inf_{\mathbf{R} \in O(3)} \ \mathbf{L}_1 - \mathbf{L}_2 \mathbf{R}\ $
where $\mathbf{A}_i = \mathbf{L}_i \mathbf{L}_i^\top$, $i = 1, 2$	

where $\inf_{\mathbf{R}} f(\mathbf{R})$ means the greatest lower bound of $f(\mathbf{R})$, with respect to \mathbf{R} . Then the Procrustes solution to estimate \mathbf{R} is

$$\hat{\mathbf{R}} = \arg \inf_{\mathbf{R} \in O(n)} \|\mathbf{L}_1 - \mathbf{L}_2 \mathbf{R}\| = \mathbf{U} \mathbf{V}^\top,$$

where $\arg \inf_{\mathbf{R}} f(\mathbf{R})$ means \mathbf{R} that gives the greatest lower bound of $f(\mathbf{R})$, $\mathbf{U}, \mathbf{V} \in O(n)$ and $\mathbf{L}_1^\top \mathbf{L}_2 = \mathbf{V} \mathbf{\Lambda} \mathbf{U}^\top$ is the singular decomposition with diagonal matrix $\mathbf{\Lambda}$ of positive singular entries (Dryden et al., 2009).

A.2 Euclidean and non-Euclidean Means

The Euclidean and non-Euclidean means for a set of covariance matrices, \mathbf{A}_i where $i \in \{1 \dots n\}$, have been summarized in Dryden et al. (2009) as follows:

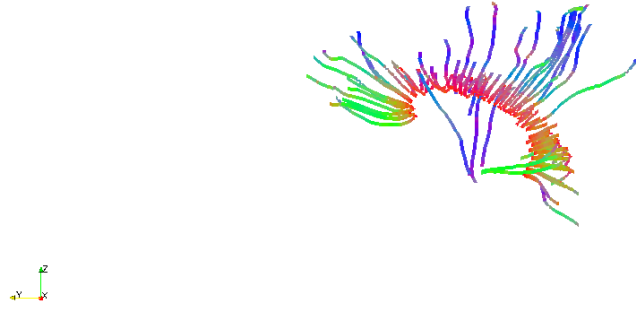
Clustering method	mean
Euclidean	$\frac{1}{n} \sum_{i=1}^n \mathbf{A}_i$
Log Euclidean	$\exp(\frac{1}{n} \sum_{i=1}^n \log \mathbf{A}_i)$
Riemannian	$\arg \inf_{\Sigma} \sum_{i=1}^n \left\ \log (\mathbf{A}_i^{-1/2} \Sigma \mathbf{A}_i^{-1/2}) \right\ ^2$
Cholesky	$(\frac{1}{n} \sum_{i=1}^n \text{chol}(\mathbf{A}_i)) (\frac{1}{n} \sum_{i=1}^n \text{chol}(\mathbf{A}_i))^\top$
Root Euclidean	$(\frac{1}{n} \sum_{i=1}^n \mathbf{A}_i^{1/2}) (\frac{1}{n} \sum_{i=1}^n \mathbf{A}_i^{1/2})^\top$
Procrustes	$\hat{\mathbf{P}} \hat{\mathbf{P}}^\top$ where $\hat{\mathbf{P}} = \arg \inf_{\mathbf{P}} \sum_{i=1}^n (\inf_{\mathbf{R}_i \in O(3)} \ \mathbf{L}_i \mathbf{R}_i - \mathbf{P}\ ^2)$ $\mathbf{A}_i = \mathbf{L}_i \mathbf{L}_i^\top$

A.3 Particular Cases of the f -mean

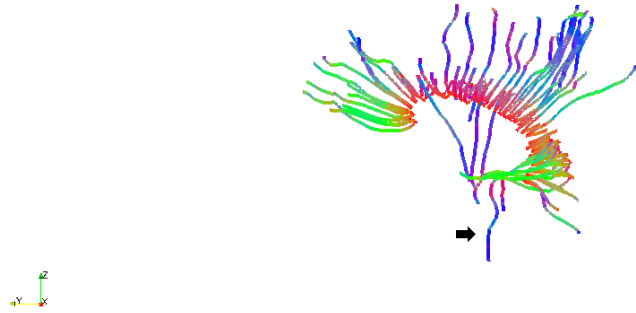
Adjusting the centroids and G_j , after the movement of a covariance matrix from cluster to another, for particular cases of the f -mean are shown below.

log Euclidean: $f(\mathbf{A}) = \log(\mathbf{A})$	
$\bar{\mathbf{A}}_{C_{np}} = \exp\left(\frac{1}{m(p)-1} [m(p) \log(\bar{\mathbf{A}}_{C_p}) - \log(\mathbf{A}_i)]\right)$	
$\bar{\mathbf{A}}_{C_{nj}} = \exp\left(\frac{1}{m(j)+1} [m(j) \log(\bar{\mathbf{A}}_{C_j}) + \log(\mathbf{A}_i)]\right)$	
$G_j = \frac{m(j)}{m(j)+1} \ \log(\mathbf{A}_i) - \log(\bar{\mathbf{A}}_{C_j})\ ^2 - \frac{m(p)}{m(p)-1} \ \log(\mathbf{A}_i) - \log(\bar{\mathbf{A}}_{C_p})\ ^2$	
root Euclidean: $f(\mathbf{A}) = \mathbf{A}^{1/2}$	
$\bar{\mathbf{A}}_{C_{np}} = \frac{1}{(m(p)-1)^2} [m(p)(\bar{\mathbf{A}}_{C_p})^{1/2} - (\mathbf{A}_i)^{1/2}] [m(p)(\bar{\mathbf{A}}_{C_p})^{1/2} - (\mathbf{A}_i)^{1/2}]^\top$	
$\bar{\mathbf{A}}_{C_{nj}} = \frac{1}{(m(j)+1)^2} [m(j)(\bar{\mathbf{A}}_{C_j})^{1/2} + (\mathbf{A}_i)^{1/2}] [m(j)(\bar{\mathbf{A}}_{C_j})^{1/2} + (\mathbf{A}_i)^{1/2}]^\top$	
$G_j = \frac{m(j)}{m(j)+1} \ (\mathbf{A}_i)^{1/2} - (\bar{\mathbf{A}}_{C_j})^{1/2}\ ^2 - \frac{m(p)}{m(p)-1} \ (\mathbf{A}_i)^{1/2} - (\bar{\mathbf{A}}_{C_p})^{1/2}\ ^2$	
Cholesky: $f(\mathbf{A}) = \text{chol}(\mathbf{A})$	
$\bar{\mathbf{A}}_{C_{np}} = \frac{1}{(m(p)-1)^2} [m(p) \text{chol}(\bar{\mathbf{A}}_{C_p}) - \text{chol}(\mathbf{A}_i)] [m(p) \text{chol}(\bar{\mathbf{A}}_{C_p}) - \text{chol}(\mathbf{A}_i)]^\top$	
$\bar{\mathbf{A}}_{C_{nj}} = \frac{1}{(m(j)+1)^2} [m(j) \text{chol}(\bar{\mathbf{A}}_{C_j}) + \text{chol}(\mathbf{A}_i)] [m(j) \text{chol}(\bar{\mathbf{A}}_{C_j}) + \text{chol}(\mathbf{A}_i)]^\top$	
$G_j = \frac{m(j)}{m(j)+1} \ \text{chol}(\mathbf{A}_i) - \text{chol}(\bar{\mathbf{A}}_{C_j})\ ^2 - \frac{m(p)}{m(p)-1} \ \text{chol}(\mathbf{A}_i) - \text{chol}(\bar{\mathbf{A}}_{C_p})\ ^2$	
Euclidean: $f(\mathbf{A}) = \mathbf{A}$	
This is Hartigan's method, which was discussed in Section 3.3.3.	

A.4 Fiber tracking

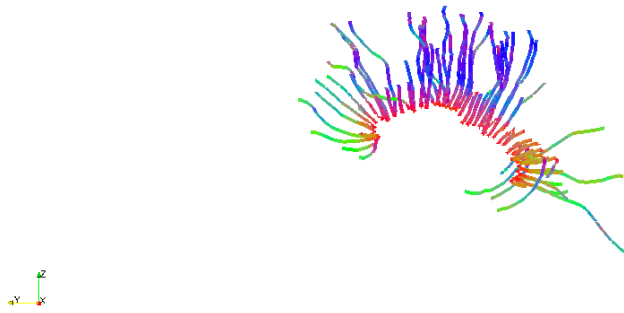


(a) Fiber tracking of the CC using the curvature threshold sc .

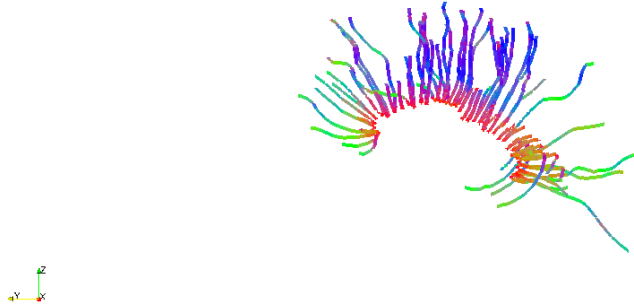


(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure A.1: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the healthy subject H2. Using 45° results in continued propagation of some fibers out of the CC (indicated with the black arrow).

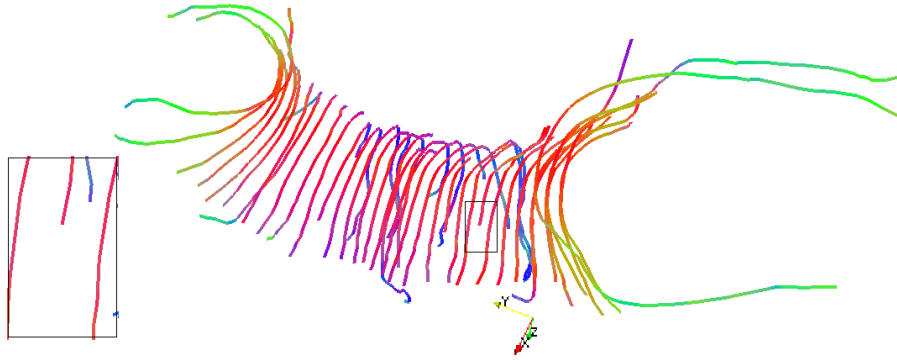


(a) Fiber tracking of the CC using the curvature threshold sc .



(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure A.2: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the healthy subject H4. Using 45° and sc provide approx. similar results.

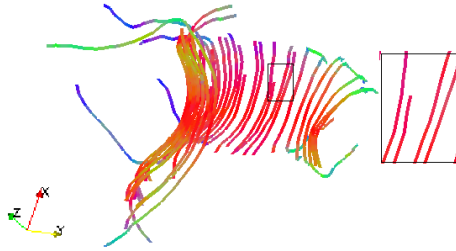


(a) Fiber tracking of the CC using the curvature threshold sc .

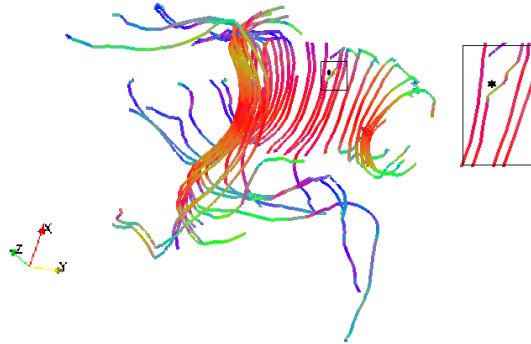


(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure A.3: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the MS subject MS2. Using 45° results in deviation of some fibers from the original pathways (indicated by the black star).

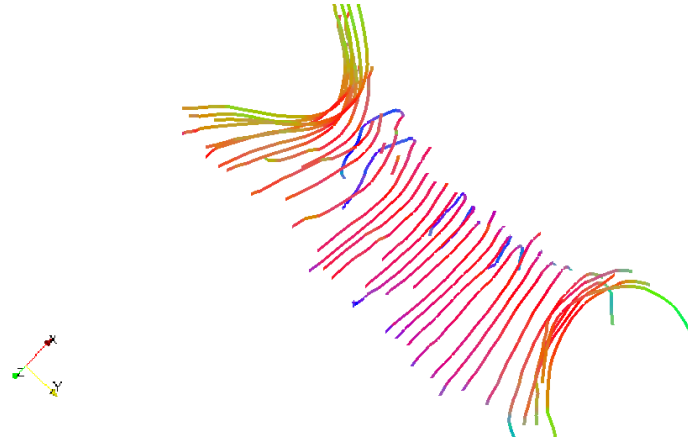


(a) Fiber tracking of the CC using the curvature threshold sc .

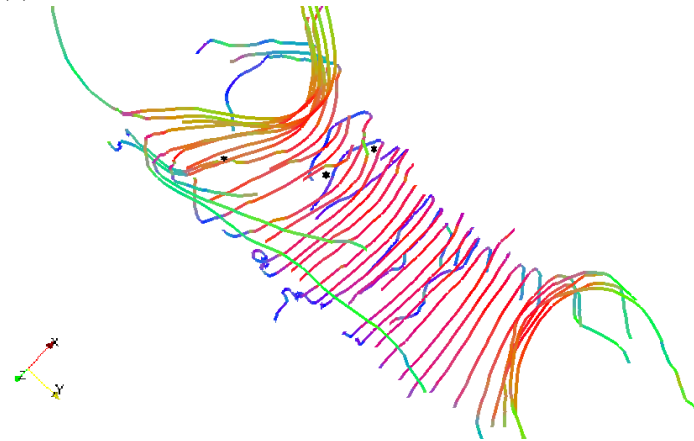


(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure A.4: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the MS subject MS3. Using 45° results in deviation of some fibers from the original pathways (indicated by the black star).

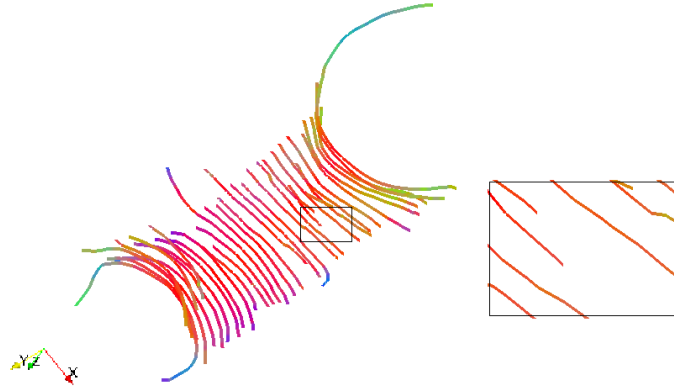


(a) Fiber tracking of the CC using the curvature threshold sc .

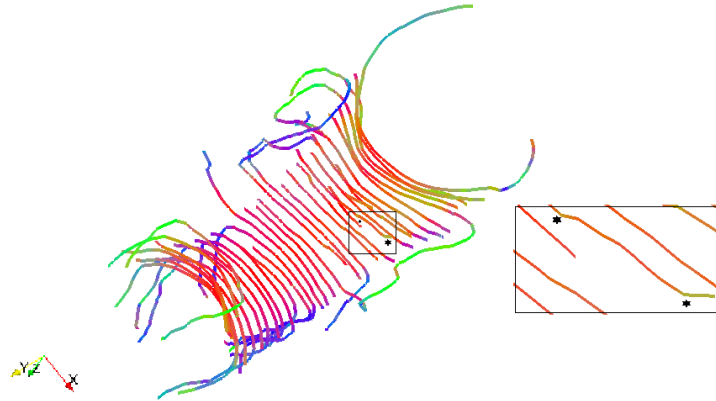


(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure A.5: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the MS subject MS4. Using 45° results in deviation of some fibers from the original pathways (indicated by the black stars).

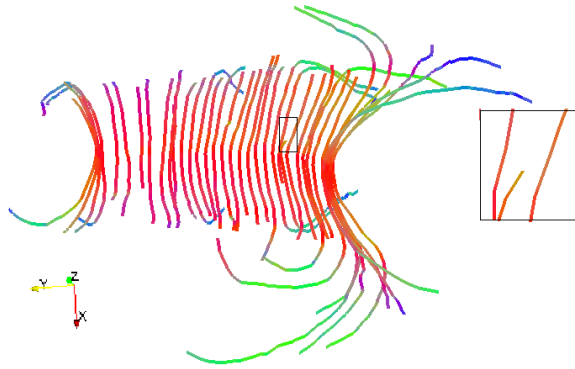


(a) Fiber tracking of the CC using the curvature threshold sc .

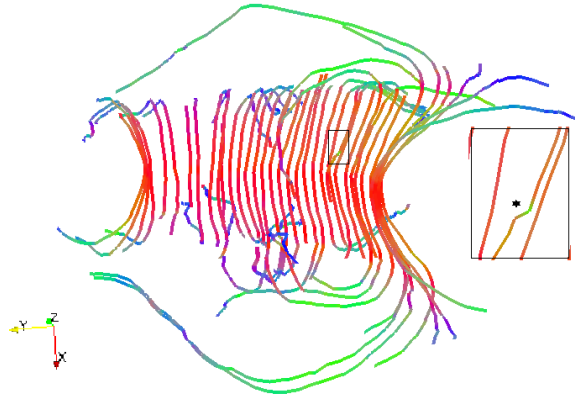


(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure A.6: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the MS subject MS5. Using 45° results in deviation of some fibers from the original pathways (indicated by the black stars).

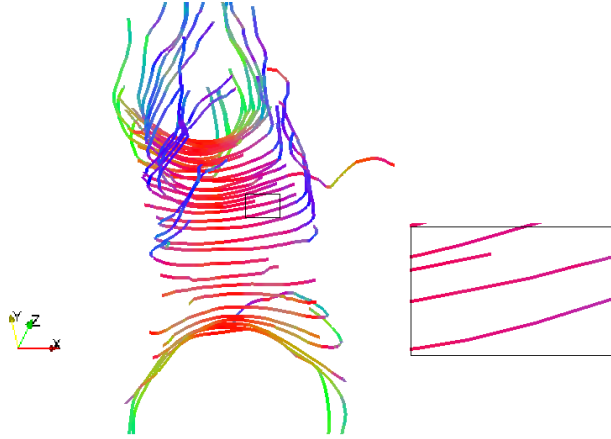


(a) Fiber tracking of the CC using the curvature threshold sc .

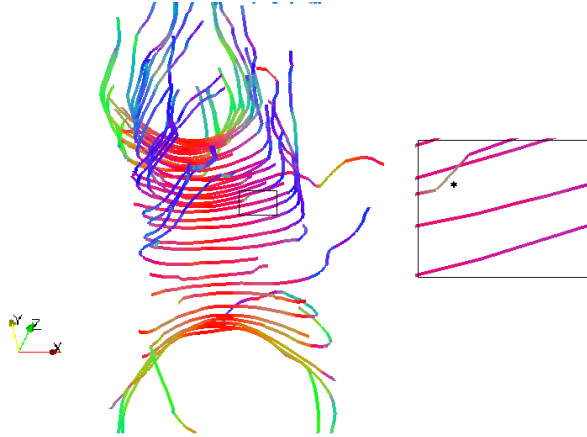


(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure A.7: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the MS subject MS7. Using 45° results in deviation of some fibers from the original pathways (indicated by the black star).

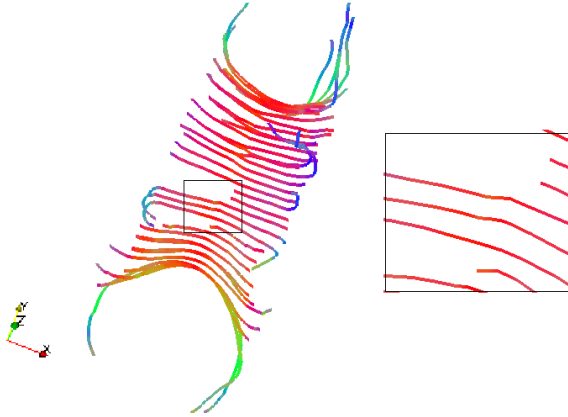


(a) Fiber tracking of the CC using the curvature threshold sc .

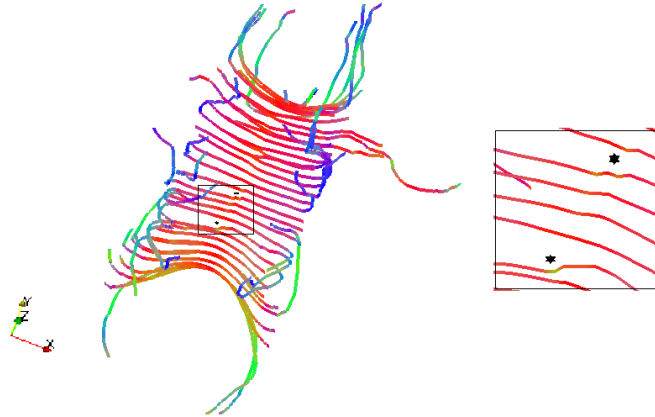


(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure A.8: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the MS subject MS8. Using 45° results in deviation of some fibers from the original pathways (indicated by the black star).



(a) Fiber tracking of the CC using the curvature threshold sc .



(b) Fiber tracking of the CC using 45° as the curvature threshold.

Figure A.9: Comparing fiber tracking of the CC using sc and the curvature threshold 45° in the MS subject MS9. Using 45° results in deviation of some fibers from the original pathways (indicated by the black stars).

References

Abramowitz, M., & Stegun, I. A. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables* (Vol. 55). Courier Corporation.

83

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle,[w:] proceedings of the 2nd international symposium on information, bn petrow, f. Czaki, *Akademiai Kiado, Budapest*.

89

Alexander, A. L., Lee, J. E., Lazar, M., Boudos, R., DuBray, M. B., Oakes, T. R., ... Lainhart, J. E. (2007). Diffusion tensor imaging of the corpus callosum in Autism. *Neuroimage*, 34(1), 61-73.

29, 33

Alexander, D. (2005). Multiple-fibre reconstruction algorithms for diffusion MRI. *Annals of the New York Academy of Sciences*, 1064, 113-133.

16

Amato, M. P., Zipoli, V., Goretti, B., Portaccio, E., De Caro, M. F., Ricchiuti, L., ... Trojano, M. (2006). Benign multiple sclerosis. *Journal of neurology*, 253(8), 1054-1059.

80

Arsigny, V., Fillard, P., Pennec, X., & Ayache, N. (2006). Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic resonance in medicine*, 56(2), 411-421.

3, 43

- Arsigny, V., Fillard, P., Pennec, X., & Ayache, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 29, 328-347.
- 35, 36, 37, 42, 77, 129
- Augen, J. (2004). *Bioinformatics in the post-genomic era: Genome, transcriptome, proteome, and information-based medicine*. Addison-Wesley Professional.
- 50
- A.V. Prokhorov, E. o. M. (2011). *Covariance matrix*. http://www.encyclopediaofmath.org/index.php?title=Covariance_matrix&oldid=13365. (Accessed 30/3/2015)
- 35
- Awate, S. P., Tasdizen, T., Foster, N., & Whitaker, R. T. (2006). Adaptive Markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification. *Medical Image Analysis*, 10(5), 726–739.
- 132
- Ay, H., Buonanno, F. S., Schaefer, P. W., Le, D. A., Wang, B., Gonzalez, R. G., & Koroshetz, W. J. (1998). Posterior leukoencephalopathy without severe hypertension utility of diffusion-weighted MRI. *Neurology*, 51(5), 1369-1376.
- 30
- Ayachit, U. (2015). The paraview guide: a parallel visualization application. 7
- Bachman, A. H., Lee, S. H., Sidtis, J. J., & Ardekani, B. A. (2014). Corpus callosum shape and size changes in early Alzheimer’s disease: a longitudinal MRI study using the OASIS brain database. *Journal of Alzheimer’s Disease*, 39(1), 71–78.
- 2, 4, 33
- Bammer, R., Augustin, M., Strasser-Fuchs, S., Seifert, T., Kapeller, P., Stollberger, R., ... Fazekas, F. (2000). Magnetic resonance diffusion tensor imaging for characterizing diffuse and focal white matter abnormalities in multiple sclerosis. *Magnetic Resonance in Medicine*, 44(4), 583–591.

- Bandyopadhyay, S., & Coyle, E. J. (2003). An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *INFOCOM 2003. twenty-second annual joint conference of the IEEE computer and communications. IEEE societies* (Vol. 3, pp. 1713–1723).
- 45
- Banerjee, A., Dhillon, I. S., Ghosh, J., & Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research*, 6(Sep), 1345–1382.
- 3, 79, 85, 86, 87, 88, 90
- Basser, P., & Jones, D. (2002). Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. *NMR Biomed*, 15(7-8), 456-67.
- 16, 33
- Basser, P., Mattiello, J., & LeBihan, D. (1994, Mar). Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B*, 103(3), 247-54.
- 6, 112
- Basser, P., Mattiello, J., & Le Bihan, D. (1994, Jan). MR diffusion tensor spectroscopy and imaging. *Biophys Journal*, 66(1), 259-67.
- 16, 17, 21, 33
- Basser, P. J. (1998). Fiber-tractography via diffusion tensor MRI (DT-MRI). In *Proceedings of the 6th annual meeting ismrm, sydney, australia* (Vol. 1226).
- 106
- Basser, P. J., 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.
- 5, 103, 106, 112, 114
- Basser, P. J., & Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance, Series B*, 111(0086), 209219.

21, 23

Behrens, T. E., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., ... Smith, S. M. (2003). Characterization and propagation of uncertainty in diffusion-weighted mr imaging. *Magnetic resonance in medicine*, 50(5), 1077–1088.

104, 110

Berger, A. (2002). Magnetic resonance imaging. *British Medical Journal*, 324(7328), 35.

12

Bhalerao, A., & Westin, C.-F. (2007). Hyperspherical von Mises-Fisher mixture (hvmf) modelling of high angular resolution diffusion MRI. In *International conference on medical image computing and computer-assisted intervention* (pp. 236–243).

3, 79

Bingham, C. (1974). An antipodally symmetric distribution on the sphere. *The Annals of Statistics*, 1201–1225.

84

Bozzali, M., Spanò, B., Parker, G., Giulietti, G., Castelli, M., Basile, B., ... others (2013). Anatomical brain connectivity can assess cognitive dysfunction in multiple sclerosis. *Multiple Sclerosis Journal*, 19(9), 1161–1168.

106

Brecheisen, R., Vilanova, A., Platel, B., & ter Haar Romeny, B. (2009). Parameter sensitivity visualization for DTI fiber tracking. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 1441–1448.

4, 104

Brown, M. A., & Semelka, R. C. (2011). *MRI: basic principles and applications*. John Wiley & Sons.

13, 14

Brown, N. M. (1994). *Exploring space shape*. <http://news.psu.edu/story/141490/1994/03/01/research/exploring-space-shape>. (Accessed

31/3/2015)

40

Budde, M. D., Kim, J. H., Liang, H.-F., Russell, J. H., Cross, A. H., & Song, S.-K. (2008). Axonal injury detected by in vivo diffusion tensor imaging correlates with neurological disability in a mouse model of multiple sclerosis. *NMR in Biomedicine*, 21(6), 589–597.

82

Bullmore, E., Brammer, M., Rouleau, G., Everitt, B., Simmons, A., Sharma, T., ... Dunn, G. (1995). Computerized brain tissue classification of magnetic resonance images: a new approach to the problem of partial volume artifact. *Neuroimage*, 2(2PA), 133–147.

132

Burzynska, A. Z., Preuschhof, C., Bäckman, L., Nyberg, L., Li, S.-C., Lindenberg, U., & Heekeren, H. R. (2010). Age-related differences in white matter microstructure: region-specific patterns of diffusivity. *Neuroimage*, 49(3), 2104–2112.

127

Callaghan, P. T. (1991). *Principles of nuclear magnetic resonance microscopy*. Clarendon Press, Oxford.

12, 13

Campbell, J. S., Siddiqi, K., Rymar, V. V., Sadikot, A. F., & Pike, G. B. (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.

104

Carmichael, O., Chen, J., Paul, D., & Peng, J. (2013). Diffusion tensor smoothing through weighted Karcher means. *Electronic Journal of Statistics*, 7, 1913–1956.

28, 35, 42

Ceccarelli, A., Rocca, M. A., Pagani, E., Ghezzi, A., Capra, R., Falini, A., ...

Filippi, M. (2008). The topographical distribution of tissue injury in benign ms: a 3t multiparametric mri study. *Neuroimage*, 39(4), 1499–1509.

80

Choi, H. S., Haynor, D. R., & Kim, Y. (1991). Partial volume tissue classification of multichannel magnetic resonance images-a mixel model. *Medical Imaging, IEEE Transactions on*, 10(3), 395–407.

132

Christidi, F., Karavasilis, E., Samiotis, K., Bisdas, S., & Papanikolaou, N. (2016). Fiber tracking: A qualitative and quantitative comparison between four different software tools on the reconstruction of major white matter tracts. *European journal of radiology open*, 3, 153–161.

104

Chung, M., Lee, J., & Alexander, A. (2005). *Anisotropic kernel smoothing in diffusion tensor imaging: Theoretical framework* (TR No. 1109). Department of Statistics, University of Wisconsin, Madison.

28

Ciccarelli, O., Werring, D. J., Barker, G. J., Griffin, C. M., Wheeler-Kingshott, C. A., Miller, D. H., & Thompson, A. J. (2003). A study of the mechanisms of normal-appearing white matter damage in multiple sclerosis using diffusion tensor imaging. *Journal of neurology*, 250(3), 287–292.

79

Cocosco, C. A., Zijdenbos, A. P., & Evans, A. C. (2003). A fully automatic and robust brain MRI tissue classification method. *Medical Image Analysis*, 7(4), 513–527.

132

Cook, P., Bai, Y., Nedjati-Gilani, S., Seunarine, K., Hall, M., Parker, G., & Alexander, D. (2006). Camino: open-source diffusion-MRI reconstruction and processing. In *14th scientific meeting of the international society for magnetic resonance in medicine* (Vol. 2759).

7, 104, 112

- Coombs, B. D., Best, A., Brown, M. S., Miller, D. E., Corboy, J., Baier, M., & Simon, J. H. (2004). Multiple sclerosis pathology in the normal and abnormal appearing white matter of the corpus callosum by diffusion tensor imaging. *Multiple Sclerosis Journal*, 10(4), 392–397.
- 82, 127
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1–38.
- 87
- Dhillon, I. S., & Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. *Machine learning*, 42(1), 143–175.
- 88
- Dhillon, I. S., & Sra, S. (2003). *Modeling data using directional distributions* (Tech. Rep.). Technical Report TR-03-06, Department of Computer Sciences, The University of Texas at Austin. URL <ftp://ftp.cs.utexas.edu/pub/techreports/tr03-06.ps.gz>.
- 87, 88
- Ding, Z., Gore, J. C., & Anderson, A. W. (2003). Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. *Magnetic Resonance in Medicine*, 49(4), 716–721.
- 132
- Dong, Q., Welsh, R. C., Chenevert, T. L., Carlos, R. C., Maly-Sundgren, P., Gomez-Hassan, D. M., & Mukherji, S. K. (2004). Clinical applications of diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 19, 618.
- 30, 31
- Downhill, J. E., Buchsbaum, M. S., Wei, T., Spiegel-Cohen, J., Hazlett, E. A., Haznedar, M. M., ... Siever, L. J. (2000). Shape and size of the corpus callosum in schizophrenia and schizotypal personality disorder. *Schizophrenia research*, 42(3), 193–208.
- 2, 33

- Dryden, I. L., Koloydenko, A. A., & Zhou, D. (2009). Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, 3(3), 1102–1123.
2, 25, 34, 35, 36, 38, 39, 40, 42, 135, 136, 137
- Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. Wiley, Chichester.
41
- Duara, R., Kushch, A., Gross-Glenn, K., Barker, W. W., Jallad, B., Pascal, S., ... others (1991). Neuroanatomic differences between dyslexic and normal readers on magnetic resonance imaging scans. *Archives of neurology*, 48(4), 410–416.
2, 33
- Dubois, J., Hertz-Pannier, L., Dehaene-Lambertz, G., Cointepas, Y., & Le Bihan, D. (2006). Assessment of the early organization and maturation of infants' cerebral white matter fiber bundles: a feasibility study using quantitative diffusion tensor imaging and tractography. *Neuroimage*, 30(4), 1121–1132.
5
- Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics* (pp. 569–593). Springer.
110
- Einstein, A. (1956). Investigations of the theory of brownian movement. *Dover, USA*. (A. D. COWPER ,Trans.)
15, 16
- El-Baz, A., Elnakib, A., Casanova, M. F., Gimelfarb, G., Switala, A. E., Jordan, D., & Rainey, S. (2011). Accurate automated detection of autism related corpus callosum abnormalities. *Journal of medical systems*, 35(5), 929–939.
2, 29, 33
- Embleton, K., Morris, D., Haroon, H., Ralph, M. L., & Parker, G. (2007). Anatomical connectivity mapping. *Proc Intl Soc Mag Reson Med*15.
105
- Farquharson, S., Tournier, J.-D., Calamante, F., Fabinyi, G., Schneider-Kolsky, M.,

Jackson, G. D., & Connelly, A. (2013). White matter fiber tractography: why we need to move beyond dti. *Journal of neurosurgery*, 118(6), 1367–1377.

5

Filippi, M., Cercignani, M., Inglese, M., Horsfield, M., & Comi, G. (2001). Diffusion tensor magnetic resonance imaging in multiple sclerosis. *Neurology*, 56(3), 304–311.

5, 82, 127

Flachaire, E. (2005). Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. *Computational Statistics & Data Analysis*, 49(2), 361–376.

110

Ford, B. J. (1992). Brownian movement in clarkia pollen: A reprise of the first observations. *The Microscope*, 40(4), 235-241.

14

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut henri poincaré* (Vol. 10, pp. 215–310).

36

Friman, O., Farneback, G., & Westin, C.-F. (2006). A bayesian approach for stochastic white matter tractography. *IEEE transactions on medical imaging*, 25(8), 965–978.

7, 110, 118

Ge, Y., Law, M., Johnson, G., Herbert, J., Babb, J. S., Mannon, L. J., & Grossman, R. I. (2004). Preferential occult injury of corpus callosum in multiple sclerosis measured by diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 20(1), 1–7.

79, 82

Girard, G., Whittingstall, K., Deriche, R., & Descoteaux, M. (2014). Towards quantitative connectivity analysis: reducing tractography biases. *Neuroimage*, 98, 266–278.

4

- Goh, A., & Vidal, R. (2008a). Clustering and dimensionality reduction on riemannian manifolds. In *Computer vision and pattern recognition, 2008. CVPR 2008. IEEE conference on* (pp. 1–7).
- 2
- Goh, A., & Vidal, R. (2008b). Segmenting fiber bundles in diffusion tensor images. In *Computer vision–eccv 2008* (pp. 238–250). Springer.
- 33, 43, 44
- Goh, A., & Vidal, R. (2008c). Unsupervised Riemannian clustering of probability density functions. In *Machine learning and knowledge discovery in databases* (pp. 377–392). Springer.
- 2
- Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, 40, 3350.
- 41
- Guo, A. C., MacFall, J. R., & Provenzale, J. M. (2002). Multiple sclerosis: diffusion tensor MR imaging for evaluation of normal-appearing white matter. *Radiology*, 222(3), 729–736.
- 79
- Gupta, R. K., Saksena, S., Hasan, K. M., Agarwal, A., Haris, M., Pandey, C. M., & Narayana, P. A. (2006). Focal wallerian degeneration of the corpus callosum in large middle cerebral artery stroke: serial diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 24(3), 549–555.
- 29, 33
- Hardan, A., Minshew, N., & Keshavan, M. (2000). Corpus callosum size in autism. *Neurology*, 55(7), 1033–1036.
- 2, 4, 29, 33
- Hartigan, J. A. (1975). *Clustering algorithms*. John Wiley & Sons, Inc.
- 2, 34, 51
- Hartlep, A., Pedain, C., Brady, M., & Raghavan, R. (2006, September 22). *Brain tissue classification*. Google Patents. (US Patent App. 11/534,383)
- 132

- Hasan, K. M., Gupta, R. K., Santos, R. M., Wolinsky, J. S., & Narayana, P. A. (2005). Diffusion tensor fractional anisotropy of the normal-appearing seven segments of the corpus callosum in healthy adults and relapsing-remitting multiple sclerosis patients. *Journal of Magnetic Resonance Imaging*, 21(6), 735–743.
- 29, 33, 105
- Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., & Tibshirani, R. (2009). *The elements of statistical learning* (Vol. 2) (No. 1). Springer.
- 45
- Heemskerk, A. M., Sinha, T. K., Wilson, K. J., Ding, Z., & Damon, B. M. (2010). Repeatability of DTI-based skeletal muscle fiber tracking. *NMR in biomedicine*, 23(3), 294–303.
- 104
- Hesseltine, S., Law, M., Babb, J., Rad, M., Lopez, S., Ge, Y., ... Grossman, R. (2006). Diffusion tensor imaging in multiple sclerosis: assessment of regional differences in the axial plane within normal-appearing cervical spinal cord. *American Journal of Neuroradiology*, 27(6), 1189–1193.
- 79
- Ho, M.-L., Moonis, G., Ginat, D. T., & Eisenberg, R. L. (2013). Lesions of the corpus callosum. *American Journal of Roentgenology*, 200(1), W1–W16.
- 65
- Hornak, J. P. (2015). *The basics of MRI*. <http://www.cis.rit.edu/htbooks/mri/>; Interactive Learning Software, Henrietta, NY. (Accessed 22/03/2015)
- 13
- Hornik, K., & Grün, B. (2014a). movMF: An R package for fitting mixtures of von Mises-Fisher distributions. *Journal of Statistical Software*, 58(10), 1–31.
- 89, 91, 92
- Hornik, K., & Grün, B. (2014b). On maximum likelihood estimation of the concentration parameter of von mises–fisher distributions. *Computational statistics*, 29(5), 945–957.

87, 92

howequipmentworks.com. (2012). http://www.howequipmentworks.com/mri_basics/. (Accessed 21/3/2015)

14

Hu, B., Ye, B., Yang, Y., Zhu, K., Kang, Z., Kuang, S., ... Shan, H. (2011). Quantitative diffusion tensor deterministic and probabilistic fiber tractography in relapsing–remitting multiple sclerosis. *European journal of radiology*, 79(1), 101–107.

104, 105

Jacob, M., Blu, T., & Unser, M. (2004). Efficient energies and algorithms for parametric snakes. *Image Processing, IEEE Transactions on*, 13(9), 1231–1244.

33

Jayasumana, S., Hartley, R., Salzmann, M., Li, H., & Harandi, M. (2013). Kernel methods on the Riemannian manifold of symmetric positive definite matrices. In *Computer vision and pattern recognition (CVPR), 2013 IEEE conference on* (pp. 73–80).

2, 43

Jbabdi, S., Woolrich, M., Andersson, J., & Behrens, T. (2007). A bayesian framework for global tractography. *Neuroimage*, 37(1), 116–129.

111

Jiang, H., Van Zijl, P. C., Kim, J., Pearlson, G. D., & Mori, S. (2006). Dtistudio: resource program for diffusion tensor computation and fiber bundle tracking. *Computer methods and programs in biomedicine*, 81(2), 106–116.

104

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3), 241–254.

45

Jones, D., & Basser, P. (2004). squashing peanuts and smashing pumpkins: How noise distorts diffusion-weighted MR data. *Magnetic Resonance in Medicine*,

52, 979993.

27

Jones, D. K. (2010). *Diffusion MRI*. Oxford University Press.

103, 107, 111

Jones, D. K., & Pierpaoli, C. (2005). Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach. *Magnetic resonance in medicine*, 53(5), 1143–1149.

110

Jones, D. K., Travis, A. R., Eden, G., Pierpaoli, C., & Basser, P. J. (2005). Pasta: pointwise assessment of streamline tractography attributes. *Magnetic Resonance in Medicine*, 53(6), 1462–1467.

110

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math*, 30, 509–541.

36

Kern, K. C., Sarcona, J., Montag, M., Giesser, B. S., & Sicotte, N. L. (2011). Corpus callosal diffusivity predicts motor impairment in relapsing–remitting multiple sclerosis: a tbss and tractography study. *Neuroimage*, 55(3), 1169–1177.

105

Klein, J., Grötsch, A., Betz, D., Barbieri, S., Friman, O., Stieltjes, B., ... Hahn, H. K. (2010). Qualitative and quantitative analysis of probabilistic and deterministic fiber tracking. In *Medical imaging: Image processing* (p. 76232A).

104

Koay, C. G., Carew, J. D., Alexander, A. L., Basser, P. J., & Meyerand, M. E. (2006). Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 55(4), 930–936.

17, 18, 19

Koch, M. A., Norris, D. G., & Hund-Georgiadis, M. (2002). An investigation

of functional and anatomical connectivity using magnetic resonance imaging. *Neuroimage*, 16(1), 241–250.

111

Koenig, K. A., Sakaie, K. E., Lowe, M. J., Lin, J., Stone, L., Bermel, R. A., ... Phillips, M. D. (2015). The relationship between cognitive function and high-resolution diffusion tensor MRI of the cingulum bundle in multiple sclerosis. *Multiple Sclerosis Journal*, 21(14), 1794–1801.

82

Kolmogorov, A. N., & Castelnovo, G. (1930). *Sur la notion de la moyenne*. G. Bardi, tip. della R. Accad. dei Lincei.

55

Kreher, B., Mader, I., & Kiselev, V. (2008). Gibbs tracking: a novel approach for the reconstruction of neuronal pathways. *Magnetic Resonance in Medicine*, 60(4), 953–963.

111

Kubicki, M., Styner, M., Bouix, S., Gerig, G., Markant, D., Smith, K., ... Shenton, M. (2008). Reduced interhemispheric connectivity in schizophrenia: tractography based segmentation of the corpus callosum. *Schizophrenia research*, 106(2), 125–131.

29, 33

Kumar, R., Bampoutis, A., Vemuri, B. C., Carney, P. R., & Mareci, T. H. (2008). Multi-fiber reconstruction from DW-MRI using a continuous mixture of von Mises-Fisher distributions. In *Computer vision and pattern recognition workshops, 2008. cvprw'08. ieee computer society conference on* (pp. 1–8).

3, 79, 85

Lachmann, F., & Barillot, C. (1992). Brain tissue classification from MRI data by means of texture analysis. In *Medical imaging VI* (pp. 72–83).

132

Lagana, M., Rovaris, M., Ceccarelli, A., Venturelli, C., Caputo, D., Cecconi, P., & Baselli, G. (2009). Atlas-based vs. individual-based deterministic tractography

of corpus callosum in multiple sclerosis. In *Engineering in medicine and biology society, 2009. embc 2009. annual international conference of the ieee* (pp. 2699–2702).

105

Lazar, M., & Alexander, A. L. (2005). Bootstrap white matter tractography (boot-trac). *NeuroImage*, 24(2), 524–532.

110

Lazar, M., Weinstein, D. M., Tsuruda, J. S., Hasan, K. M., Arfanakis, K., Meyerand, M. E., ... others (2003). White matter tractography using diffusion tensor deflection. *Human brain mapping*, 18(4), 306–321.

105, 107, 108, 109

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, 401–407.

15

Le Bihan, D., Mangin, J., Poupon, C., Clark, C., Pappata, S., Molko, N., & Chabria, H. (2001). Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4), 534–46.

16, 21, 31

Lee, J., Chung, M., Oakes, T., & Alexander, A. (2005). Anisotropic gaussian kernel smoothing of DTI data. In *Proceedings of the 13th annual meeting of ISMRM* (p. 2253).

28

Lee, S.-P., Cheng, J.-Z., Chen, C.-M., & Tseng, W.-Y. I. (2008). An automatic segmentation approach for boundary delineation of corpus callosum based on cell competition. In *Engineering in medicine and biology society, 2008. embs 2008. 30th annual international conference of the IEEE* (pp. 5514–5517).

33

Lefevre, J., Auzias, G., & Germanaud, D. (2014). Brain lobes revealed by spectral clustering. In *Pattern recognition (icpr), 2014 22nd international conference*

on (pp. 562–567).

132

Lenglet, C., Rousson, M., & Deriche, R. (2006). A statistical framework for DTI segmentation. In *Biomedical imaging: Nano to macro, 2006. 3rd IEEE international symposium on* (pp. 794–797).

2, 3, 33, 43, 131

Li, Y., Jewells, V., Kim, M., Chen, Y., Moon, A., Armao, D., . . . Shen, D. (2013). Diffusion tensor imaging based network analysis detects alterations of neuroconnectivity in patients with clinically early relapsing-remitting multiple sclerosis. *Human brain mapping*, 34(12), 3376–3391.

104

Li, Y., Mandal, M., & Ahmed, S. N. (2013). Fully automated segmentation of corpus callosum in midsagittal brain MRIs. In *Engineering in medicine and biology society (EMBC), 2013 35th annual international conference of the IEEE* (pp. 5111–5114).

33

Liu, R. Y., et al. (1988). Bootstrap procedures under some non-iid models. *The Annals of Statistics*, 16(4), 1696–1708.

110

Lloyd, S. P. (1982). Least squares quantization in PCM. *Information Theory, IEEE Transactions on*, 28(2), 129–137.

2, 34

Lundervold, A., Duta, N., Taxt, T., & Jain, A. K. (1999). Model-guided segmentation of corpus callosum in MR images. In *Computer vision and pattern recognition, 1999. IEEE computer society conference on*. (Vol. 1).

33

Lyksborg, M., Siebner, H. R., Sørensen, P. S., Blinkenberg, M., Parker, G. J., Dogonowski, A.-M., . . . Dyrby, T. B. (2014). Secondary progressive and relapsing remitting multiple sclerosis leads to motor-related decreased anatomical connectivity. *PloS one*, 9(4), e95540.

104, 105

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297).

50

Makki, M., Chugani, D., Janisse, J., & Chugani, H. (2007). Characteristics of abnormal diffusivity in normal-appearing white matter investigated with diffusion tensor MR imaging in tuberous sclerosis complex. *American Journal of Neuroradiology*, 28(9), 1662–1667.

5

Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. *J. Phys. C: Solid State Phys*, 10, L55-L58.

15

Marcella Laganà, M., Ceccarelli, A., Giulia Preti, M., Venturelli, C., Pia Sormani, M., Cavarretta, R., ... Rovaris, M. (2012). Atlas-based versus individual-based fiber tracking of the corpus callosum in patients with multiple sclerosis: Reliability and clinical correlations. *Journal of Neuroimaging*, 22(4), 355–364.

104

Mardia, K. V., & Jupp, P. E. (2009). *Directional statistics* (Vol. 494). John Wiley & Sons.

3, 79, 83

McGraw, T., Vemuri, B., Yezierski, R., & Mareci, T. (2006). Segmentation of high angular resolution diffusion MRI modeled as a field of von Mises-Fisher mixtures. In *European conference on computer vision* (pp. 463–475).

3, 79

Mesaros, S., Rocca, M. A., Riccitelli, G., Pagani, E., Rovaris, M., Caputo, D., ... others (2009). Corpus callosum damage and cognitive dysfunction in benign ms. *Human brain mapping*, 30(8), 2656–2666.

105

Miller, D. H., & Leary, S. M. (2007). Primary-progressive multiple sclerosis. *The*

Lancet Neurology, 6(10), 903–912.

80

Miller, S. P., Vigneron, D. B., Henry, R. G., Bohland, M. A., Ceppi-Cozzio, C., Hoffman, ., C., & Barkovich, A. J. (2002). Serial quantitative diffusion tensor MRI of the premature brain: development in newborns with and without injury. *Journal of Magnetic Resonance Imaging*, 16(6), 621-632.

29

Minitab, I. (2010). Minitab 17 statistical software state college pa. 6

Morgan, M. A., & Jones, J. (2015a). *T1 weighted image*. <http://radiopaedia.org/articles/t1-weighted-image>. (Accessed 20/3/2015)

14

Morgan, M. A., & Jones, J. (2015b). *T2 weighted image*. <http://radiopaedia.org/articles/t2-weighted-image>. (Accessed 20/3/2015)

14

Mori, S., Crain, B. J., Chacko, V. P., & Van Zijl, P. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of neurology*, 45(2), 265–269.

107

Mori, S., Wakana, S., Nage-Poetscher, L., & Van Zijl, P. (2005). *MRI atlas of human white matter* (T. N. Amsterdam, Ed.). Elsevier.

23

Morris, D. M., Embleton, K. V., & Parker, G. J. (2008). Probabilistic fibre tracking: differentiation of connections from chance events. *Neuroimage*, 42(4), 1329–1339.

103

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1), 90–100.

85, 86

Nazem-Zadeh, M.-R., Saksena, S., Babajani-Fermi, A., Jiang, Q., Soltanian-Zadeh,

H., Rosenblum, M., ... Jain, R. (2012). Segmentation of corpus callosum using diffusion tensor imaging: validation in patients with glioblastoma. *BMC medical imaging*, 12(1), 10.

2, 33, 34

O'Dwyer, R., Wehner, T., LaPresto, E., Ping, L., Tkach, J., Noachtar, S., & Diehl, B. (2010). Differences in corpus callosum volume and diffusivity between temporal and frontal lobe epilepsy. *Epilepsy & Behavior*, 19(3), 376–382.

29, 33

Oh, J., Henry, R., Genain, C., Nelson, S., & Pelletier, D. (2004). Mechanisms of normal appearing corpus callosum injury related to pericallosal t1 lesions in multiple sclerosis using directional diffusion tensor and 1 h mrs imaging. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(9), 1281–1286.

79

Ozturk, A., Smith, S., Gordon-Lipkin, E., Harrison, D., Shiee, N., Pham, D., ... Reich, D. (2010). MRI of the corpus callosum in multiple sclerosis: association with disability. *Multiple Sclerosis Journal*, 16(2), 166–177.

104

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.

5

Painter, K., & Hillen, T. (2013). Mathematical modelling of glioma growth: the use of diffusion tensor imaging (DTI) data to predict the anisotropic pathways of cancer invasion. *Journal of theoretical biology*, 323, 25–39.

80

Palm Dzung, L., Chenyang, X., & Prince Jerry, L. (1998). A survey of current methods in medical image segmentation..

132

Parker, G. J., & Alexander, D. C. (2003). Probabilistic monte carlo based mapping

of cerebral connections utilising whole-brain crossing fibre information. In *Ipmi* (Vol. 18, pp. 684–695).

110

Parker, G. J., Haroon, H. A., & Wheeler-Kingshott, C. A. (2003). A framework for a streamline-based probabilistic index of connectivity (pico) using a structural interpretation of mri diffusion measurements. *Journal of Magnetic Resonance Imaging*, 18(2), 242–254.

111

Parker, G. J., Wheeler-Kingshott, C. A., & Barker, G. J. (2002). Estimating distributed anatomical connectivity using fast marching methods and diffusion tensor imaging. *IEEE transactions on medical imaging*, 21(5), 505–512.

111

Pennec, X. (1991). Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements. in cetin, a., akarun, l., ertuzun, a., gurcan, m., and yadimci, y. *Proc. of Nonlinear Signal and Image Processing (NSIP99)*, 1, 194198. (June 20-23, Antalya, Turkey. IEEE-EURASIP)

41

Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision, Springer Verlag (Germany)*, 25(1), 127-154.

35, 41

Pennec, X., Fillard, P., & Ayache, N. (2006). A Riemannian framework for tensor computing. *International Journal of Computer Vision (Springer)*, 66(1), 41-66.

3, 35, 37

Pierpaoli, C., & Basser, P. J. (1996). Toward a quantitative assessment of diffusion anisotropy. *Magnetic resonance in Medicine*, 36(6), 893-906.

24

Pigoli, D., Aston, J. A., Dryden, I. L., & Secchi, P. (2014). Distances and inference for covariance operators. *Biometrika*, 101(2), 409-422.

41

Pine, A., Jones, S., Lowe, M., Sakaie, K., & Phillips, M. (2009). Fiber-tracking through multiple sclerosis lesions using probabilistic tracking. In *17th annual meeting of the international society for magnetic resonance in medicine* (pp. 18–24).

104

Preziosa, P., Rocca, M. A., Mesaros, S., Pagani, E., Stosic-Opincal, T., Kacar, K., ... others (2011). Intrinsic damage to the major white matter tracts in patients with different clinical phenotypes of multiple sclerosis: a voxelwise diffusion-tensor mr study. *Radiology*, *260*(2), 541–550.

104

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/>

7

Ratnay, M. (2000). A model-based distance for clustering. In *Neural networks, 2000. IJCNN 2000, proceedings of the IEEE-INNS-ENNS international joint conference on* (Vol. 4, pp. 13–16).

2

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, *297*(5586), 1551–1555.

45

Reynolds, G., Nir, T. M., Jahanshad, N., Prasad, G., & Thompson, P. M. (2014). Using the raw diffusion MRI signal and the von Mises-Fisher distribution for classification of Alzheimer’s disease. In *Biomedical imaging (isbi), 2014 ieee 11th international symposium on* (pp. 1027–1030).

3, 80

Rockafellar, R. T. (1993). Lagrange multipliers and optimality. *SIAM review*, *35*(2), 183–238.

Rosas, H. D., Lee, S. Y., Bender, A. C., Zaleta, A. K., Vangel, M., Yu, P., . . . Hersch, S. M. (2010). Altered white matter microstructure in the corpus callosum in huntington's disease: implications for cortical disconnection. *Neuroimage*, 49(4), 2995–3004.

29, 33

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53–65.

68

Rovaris, M., Iannucci, G., Falautano, M., Possa, F., Martinelli, V., Comi, G., & Filippi, M. (2002). Cognitive dysfunction in patients with mildly disabling relapsing–remitting multiple sclerosis: an exploratory study with diffusion tensor MR imaging. *Journal of the neurological sciences*, 195(2), 103–109.

79

Sandhu, R., Georgiou, T., & Tannenbaum, A. (2008). A new distribution metric for image segmentation. In *Medical imaging* (pp. 691404–691404).

33

Sanjay-Gopal, S., & Hebert, T. J. (1998). Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. *Image Processing, IEEE Transactions on*, 7(7), 1014–1028.

132

Scheel, M., Finke, C., Oberwahrenbrock, T., Freing, A., Pech, L.-M., Schlichting, J., . . . Brandt, A. U. (2014). Retinal nerve fibre layer thickness correlates with brain white matter damage in multiple sclerosis: a combined optical coherence tomography and diffusion tensor imaging study. *Multiple Sclerosis Journal*, 20(14), 1904–1907.

82

Schwarz, G., et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461–464.

- Seber, G. A. (2009). *Multivariate observations* (Vol. 252). John Wiley & Sons.
34, 46
- Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A., & Leahy, R. M. (2001). Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, *13*(5), 856–876.
132
- Shu, N., Liu, Y., Li, K., Duan, Y., Wang, J., Yu, C., . . . He, Y. (2011). Diffusion tensor tractography reveals disrupted topological efficiency in white matter structural networks in multiple sclerosis. *Cerebral Cortex*, *21*(11), 2565–2577.
104
- Smith, R. E., Tournier, J.-D., Calamante, F., & Connelly, A. (2012). Anatomically-constrained tractography: improved diffusion mri streamlines tractography through effective use of anatomical information. *Neuroimage*, *62*(3), 1924–1938.
4
- Song, H., Liu, J., & Wang, G. (2012). High-order parameter approximation for von mises–fisher distributions. *Applied Mathematics and Computation*, *218*(24), 11880–11890.
87
- 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.
5, 82
- Spath, H. (1985). *The cluster dissection and analysis theory FORTRAN programs examples*. Prentice-Hall, Inc.
54
- Sra, S. (2012). A short note on parameter approximation for von mises-fisher distributions: and a fast implementation of $i s(x)$. *Computational Statistics*, *27*(1), 177–190.

Stejskal, E. O., & Tanner. (1965). Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *Journal of Chemical Physics*, 42, 288-292.

Suckling, J., Sigmundsson, T., Greenwood, K., & Bullmore, E. (1999). A modified fuzzy clustering algorithm for operator independent brain tissue classification of dual echo MR images. *Magnetic resonance imaging*, 17(7), 1065–1076.

Sullivan, E. V., Adalsteinsson, E., & Pfefferbaum, A. (2005). Selective age-related degradation of anterior callosal fiber bundles quantified in vivo with fiber tracking. *Cerebral Cortex*, 16(7), 1030–1039.

Sullivan, E. V., Rohlfing, T., & Pfefferbaum, A. (2010). Quantitative fiber tracking of lateral and interhemispheric white matter systems in normal aging: relations to timed performance. *Neurobiology of aging*, 31(3), 464–481.

Suri, J. S. (2004, August 31). *Segmentation method and apparatus for medical images using diffusion propagation, pixel classification, and mathematical morphology*. (US Patent 6,785,409)

Tabelow, K., Polzehl, J., Spokoiny, V., & Voss, H. U. (2008). Diffusion tensor imaging: Structural adaptive smoothing. *NeuroImage*, 39, 1763-1773.

Tan, P.-N., Steinbach, M., Kumar, V., et al. (2006). *Introduction to data mining* (Vol. 1). Pearson Addison Wesley Boston.

Tanabe, A., Fukumizu, K., Oba, S., Takenouchi, T., & Ishii, S. (2007). Parameter estimation for von mises–fisher distributions. *Computational Statistics*, 22(1), 145–157.

87

teaching.shu.ac.uk. (2015). *Nuclear magnetic resonance spectroscopy*. <http://teaching.shu.ac.uk/hwb/chemistry/tutorials/molspec/nmr1.htm>. (Accessed 21/3/2015)

13

Telgarsky, M., & Vattani, A. (2010). Hartigan's method: k-means clustering without voronoi. In *International conference on artificial intelligence and statistics* (p. 820-827).

2, 34

Tievsky, A. L., Ptak, T., & Farkas, J. (1999). Investigation of apparent diffusion coefficient and diffusion tensor anisotropy in acute and chronic multiple sclerosis lesions. *American Journal of Neuroradiology*, 20(8), 1491–1499.

82, 127

Tikhomirov, V. M. (2012). *Selected works of an kolmogorov: Volume i: Mathematics and mechanics* (Vol. 25). Springer Science & Business Media.

55

Tovar-Moll, F., Evangelou, I., Chiu, A., Richert, N., Ostuni, J., Ohayon, J., ... others (2009). Thalamic involvement and its impact on clinical disability in patients with multiple sclerosis: a diffusion tensor imaging study at 3t. *American Journal of Neuroradiology*, 30(7), 1380–1386.

82

Uk.mathworks.com. (2015a). *Agglomerative hierarchical cluster tree - matlab linkage*. Retrieved from <http://uk.mathworks.com/help/stats/linkage.html> (Accessed 1 March 2015)

47, 64

Uk.mathworks.com. (2015b). *Construct agglomerative clusters from linkages - matlab cluster*. Retrieved from <http://uk.mathworks.com/help/stats/cluster.html> (Accessed 1 March 2015)

64

Van Leemput, K., Maes, F., Vandermeulen, D., & Suetens, P. (1999). Automated

model-based bias field correction of MR images of the brain. *Medical Imaging, IEEE Transactions on*, 18(10), 885–896.

132

Wahl, M., Hübers, A., Lauterbach-Soon, B., Hattingen, E., Jung, P., Cohen, L. G., & Ziemann, U. (2011). Motor callosal disconnection in early relapsing-remitting multiple sclerosis. *Human brain mapping*, 32(6), 846–855.

105

Walterfang, M., Fahey, M., Abel, L., Fietz, M., Wood, A., Bowman, E., . . . Velakoulis, D. (2011). Size and shape of the corpus callosum in adult Niemann-Pick type C reflects state and trait illness variables. *American Journal of Neuroradiology*, 32(7), 1340–1346.

2, 33

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. *IEEE Trans. Med. Imaging*, 23, 930939.

39, 42

Watson, G. (1965). Equatorial distributions on a sphere. *Biometrika*, 52(1/2), 193–201.

84

Weinstein, D., Kindlmann, G., & Lundberg, E. (1999). Tensorlines: Advection-diffusion based propagation through diffusion tensor fields. In *Proceedings of the conference on visualization'99: celebrating ten years* (pp. 249–253).

108

Werring, D. J., Clark, C. A., Barker, G. J., Thompson, A. J., & Miller, D. H. (1999). Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology*, 52(8), 1626–1626.

30, 79, 82

Whitcher, B., Tuch, D. S., Wisco, J. J., Sorensen, A. G., & Wang, L. (2008). Using the wild bootstrap to quantify uncertainty in diffusion tensor imaging. *Human brain mapping*, 29(3), 346–362.

7, 110, 118

Witelson, S. F. (1989). Hand and sex differences in the isthmus and genu of the human corpus callosum. *Brain*, 112(3), 799–835.

101

Wood, A. T. (1994). Simulation of the von mises fisher distribution. *Communications in statistics-simulation and computation*, 23(1), 157–164.

90

Xu, T., Mandal, M., Long, R., Cheng, I., & Basu, A. (2012). An edge-region force guided active shape approach for automatic lung field detection in chest radiographs. *Computerized Medical Imaging and Graphics*, 36(6), 452–463.

70

Yaldizli, Ö., Pardini, M., Sethi, V., Muhlert, N., Liu, Z., Tozer, D. J., ... others (2016). Characteristics of lesional and extra-lesional cortical grey matter in relapsing–remitting and secondary progressive multiple sclerosis: A magnetisation transfer and diffusion tensor imaging study. *Multiple Sclerosis Journal*, 22(2), 150–159.

83

Yang, F., Zhu, Y.-M., Magnin, I. E., Luo, J.-H., Croisille, P., & Kingsley, P. B. (2012, February). Feature-based interpolation of diffusion tensor fields and application to human cardiac DT-MRI. *Medical Image Analysis*, 16(2), 459–481.

42

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*, 31(3), 1116–1128.

7

Zahr, N. M., Rohlfing, T., Pfefferbaum, A., & Sullivan, E. V. (2009). Problem solving, working memory, and motor correlates of association and commissural fiber bundles in normal aging: a quantitative fiber tracking study. *Neuroimage*, 44(3), 1050–1062.

127

Zhang, F., Hancock, E. R., Goodlett, C., & Gerig, G. (2009). Probabilistic white matter fiber tracking using particle filtering and von Mises-Fisher sampling. *Medical image analysis*, 13(1), 5–18.

80

Zhou, D., Dryden, I. L., Koloydenko, A., & Bai, L. (2013). Procrustes analysis for diffusion tensor image processing. *International Journal of Computer Theory and Engineering*, 5(1), 108–113.

25, 26

Zhou, D., Dryden, I. L., Koloydenko, A. A., Audenaert, K. M., & Bai, L. (2016). Regularisation, interpolation and visualisation of diffusion tensor images using non-Euclidean statistics. *Journal of Applied Statistics*, 43(5), 943–978.

35, 41, 42, 77