

Quantitative Analysis of Cerebral White Matter Anatomy from Diffusion MRI

by

Mahnaz Maddah

M.Sc., Electrical Engineering, University of Tehran, 2002

B.Sc., Electrical Engineering, University of Tehran, 2000

Submitted to the

Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2008

© Massachusetts Institute of Technology 2008. All rights reserved.

Author
Department of Electrical Engineering and Computer Science
August 14, 2008

Certified by
W. Eric L. Grimson
Bernard Gordon Professor of Medical Engineering
Thesis Supervisor

Certified by
William M. Wells
Associate Professor of Radiology, Harvard Medical School
Thesis Supervisor

Certified by
Carl-Fredrik Westin
Associate Professor of Radiology, Harvard Medical School
Thesis Supervisor

Accepted by
Terry P. Orlando
Professor of Electrical Engineering
Chairman, Department Committee on Graduate Students

Quantitative Analysis of Cerebral White Matter Anatomy from Diffusion MRI

by

Mahnaz Maddah

Submitted to the Department of Electrical Engineering and Computer Science
on August 14, 2008, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

In this thesis we develop algorithms for quantitative analysis of white matter fiber tracts from diffusion MRI. The presented methods enable us to look at the variation of a diffusion measure along a fiber tract in a single subject or a population, which allows important clinical studies toward understanding the relation between the changes in the diffusion measures and brain diseases, development, and aging. The proposed quantitative analysis is performed on a group of fiber trajectories extracted from diffusion MRI by a process called tractography. To enable the quantitative analysis we first need to cluster similar trajectories into groups that correspond to anatomical bundles and to establish the point correspondence between these variable-length trajectories. We propose a computationally-efficient approach to find the point correspondence and the distance between each trajectory to the prototype center of each bundle. Based on the computed distances we also develop a novel model-based clustering of trajectories into anatomically-known fiber bundles. In order to cluster the trajectories, we formulate an expectation maximization algorithm to infer the parameters of the gamma-mixture model that we built on the distances between trajectories and cluster centers. We also extend the proposed clustering algorithm to incorporate spatial anatomical information at different levels through hierarchical Bayesian modeling. We demonstrate the effectiveness of the proposed methods in several clinical applications. In particular, we present our findings in identifying localized group differences in fiber tracts between normal and schizophrenic populations.

Thesis Supervisor: W. Eric L. Grimson

Title: Bernard Gordon Professor of Medical Engineering

Thesis Supervisor: William M. Wells

Title: Associate Professor of Radiology, Harvard Medical School

Thesis Supervisor: Carl-Fredrik Westin

Title: Associate Professor of Radiology, Harvard Medical School

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Prof. Eric Grimson. During the past four years, I never left his office without feeling more inspired and more motivated about my research. His vivid insight to the field and his constant support were what got me through during the difficult stages of my Ph.D. course.

I learned a lot from my co-supervisor, Prof. Sandy Wells; I found it almost impossible for a mathematical bug to be missed by him. Thanks Sandy! Dr. C-F Westin's knowledge on diffusion MRI has been a treasure for me. I thank him for his help and for introducing me to Dr. Shenton's group where I had the opportunity to meet and work with very knowledgeable people, in particular Dr. Marek Kubicki. I would also like to thank Dr. Ron Kikinis, who accepted me very warmly in the first week I came to the United States in 2003 and introduced me to the very distinguished people whom I had the honor to work with from then on. I had the pleasure to be in Simon Warfield's group and get to know a lot about the field of medical image analysis. I'm thankful to my Ph.D. committee members, Prof. Alan Willsky and Prof. Polina Golland for reading my thesis and providing me with constructive comments and insightful feedback.

I would like to thank my colleagues over SPL, CRL, PNL and LMI, especially, Andrea Mewes, Steven Haker, and my wonderful officemates Kilian Pohl, Lauren O'Donnell and Ulas Ziyan for their collaboration and helpful discussions. My special thanks are dedicated to Lilla Zöllei for providing me with her fabulous congealing code, for her help on Bayesian stuff, and for careful reading of my thesis and papers. I would also like to thank Eric's assistants Donna Kaufman, and Maysoon Hamdin.

My life in Boston started with meeting two of my husband's friends, Mohsen Razavi and Hazhir Rahmandad, whom since then have constantly raised my admiration for their superb talents and their kind and genuine spirit. I also like to thank my dearest Sanaz; her smiley face and her constant appreciation of whatever I do have always been heart-warming for me. So many thanks to my dear friends, Mana, Mehdi, Melanie, Rosa, Hamed, Ghazal, Niayesh, Shaya, Mohammad, Leila, Payam, Saeed,

Maryam, Fardad, Sara, Maryam, Arash, Afsaneh, Hossein, Nicole, and Mohammad who made the Bostonian life fun and manageable while being apart from all the family and relatives. By all means, this acknowledgment would not be complete without thanking a good friend and colleague of mine, Ali Shahrokni, whom inspired me to switch from electronics to the field of medical image analysis; an area of research that became my professional passion.

Above all, I owe my achievements to my parents for their belief in me, from the earliest time that I can remember, and I certainly would not reach this point without their constant support and encouragement to continue my studies. “maman, baba merci!!!”. I would like to thank my sisters, Mahshid and Mahdokht and my brother, Mahdi, who have always been simply the best friends one could have ever wished for. Thanks to my parents-in-law also for including me in their good prayers.

I would have not been able to finish this Ph.D. program without the help, support and love I received from my dear husband, Ali Khakifrooz. I take the chance here to thank him from the bottom of my heart. The shining of my little star, Setareh, in our life made this period more challenging for me, but also more fun and joyful. Thank you my sweet love ...

Contents

1	Introduction	23
1.1	Quantitative Analysis of Diffusion MRI and Its Challenges	25
1.2	Objective and Contributions of The Thesis	30
1.3	Organization of the Thesis	34
2	Background on Diffusion MRI Quantitative Analysis	37
2.1	Diffusion Anisotropy in White Matter	38
2.2	Diffusion Weighted MRI	40
2.3	Diffusion Tensor MRI	41
2.4	Tensor Visualization and Scalar Measures of Diffusion	43
2.4.1	Mean Diffusivity	44
2.4.2	Anisotropy Measures	45
2.4.3	Orientation Mapping	45
2.4.4	Geometrical Anisotropy Measures	46
2.5	Tractography	47
2.5.1	Streamline Tractography	49
2.5.2	Diffusion Deflection Tractography	49
2.5.3	Stochastic Tractography	50
2.6	Related Work on Quantitative Analysis of White Matter Fiber Tracts	51
3	Point Correspondence and Similarity Measures for 3-D Curves	57
3.1	Similarity Measures for Curves	58
3.1.1	Bottleneck Distance	58

3.1.2	Hausdorff Distance	58
3.1.3	Fréchet Distance	59
3.1.4	Nonlinear Elastic Matching Distance	59
3.1.5	Turning Function Distance	59
3.1.6	Reflection Distance	60
3.1.7	Transport Distance	60
3.2	Similarity Measures Used in DTI Analysis	61
3.3	Our Proposed Approach: Distance Transforms	62
3.4	Discussion	66
3.5	Summary	67
4	Probabilistic Clustering of White Matter Trajectories	71
4.1	Desired Features and Possible Solutions	72
4.2	Probabilistic Model for Trajectories	74
4.2.1	Density Estimation	74
4.2.2	Mixture Model	76
4.3	Graphical Model for Our Clustering Problem	78
4.4	Expectation-Maximization Algorithm	79
4.5	Implementation of Our Proposed Clustering Approach	82
4.5.1	Application of EM to clustering of white matter trajectories .	82
4.5.2	Updating the Cluster Centers and Distance Matrix	87
4.5.3	Initialization	88
4.5.4	Handling Outliers	89
4.6	Results	90
4.6.1	Impact of Initial Cluster Centers	90
4.6.2	Outlier Rejection	91
4.6.3	Whole Brain Analysis and Population Studies	92
4.7	Summary	96
5	Integration of Anatomical Information into Clustering of White Mat- ter Trajectories	101

5.1	Representation of Anatomical Knowledge	102
5.2	Related Work	104
5.3	Membership Assignments from the Atlas	105
5.4	Fixed-Weight Prior (FWP)	107
5.4.1	EM Formulation for FWP	107
5.5	Adjustable-Weight Prior (AWP)	110
5.5.1	Hierarchical Bayes and Conjugate Prior	112
5.5.2	Basic Definitions for AWP Modeling	114
5.5.3	EM Formulation for AWP	117
5.5.4	Results	122
5.6	Registration	125
5.7	Summary	127
6	Experimental Results on Clinical Data	131
6.1	Three-Dimensional Modeling and Visualization of Bundles	131
6.2	Investigation of Neonate Brain Development	134
6.2.1	Material and Method	135
6.2.2	Results	135
6.3	DT-MRI Findings in Schizophrenia	136
6.3.1	Method	138
6.3.2	Experiments and Results	139
6.3.3	Discussion	150
6.4	Summary	154
7	Conclusions	157
A	Graphical Models	159

List of Figures

1-1	(a) A sagittal dissection of postmortem brain that shows white matter fiber tracts (From Virtual Hospital [98]). (b) An example of fiber trajectories reconstructed from diffusion tensor images.	24
1-2	An example of tract-oriented quantitative analysis performed on one part of the corpus callosum of a subject diagnosed with multiple sclerosis. (a) A group of the fiber trajectories are colored based on the local fractional anisotropy (FA). (b) FA variation along the trajectories (blue) along with mean (solid red), and mean plus/minus standard deviation (dashed red) of the FA. The tract-oriented analysis identifies the FA reduction in the vicinity of the lesion without prior knowledge of where the lesion is located.	29
1-3	An example of tract-oriented quantitative analysis on a population and for the splenium parts of the corpus collasum. (a) Sagittal view of the trajectories from all cases, registered into the atlas space, and colored with the local fractional anisotropy. (b) Sagittal view of the trajectories in (a) that are clustered into the upper and lower parts of the splenium. (c) Axial view of the trajectories shown in (b). (d) Average fractional anisotropy along the tract arc length for the normal (blue) and schizophrenia (red) cases, where a drop in FA is seen for diseased cases. (e) A plot of the corresponding p-values which indicates that a significant drop in the FA is only observed in the mid-portion of each side of the splenium.	31

1-4	(a) Trajectories of 5 different clusters: splenium (yellow), corticospinal (red), corticobulbar (green), middle cerebellar peduncle (blue), and genu (magenta). (b) A model representation of the bundles as the mean trajectory and the isosurfaces corresponding to spatial variation of the clusters.	34
2-1	A schematic view of a myelinated axon [12]. Myelin, the axonal membrane, microtubules, and neurofilaments are all oriented in the direction of the axon and could restrict water diffusion. Hence, the apparent diffusion coefficient perpendicular to the axon, $D(\perp)$, will be smaller than the one parallel to the axon, $D(\parallel)$	40
2-2	Samples of diffusion weighted images acquired from human brain. The first two images are baseline images taken with small b -factor, while the rest are diffusion weighted images. The gradient direction is different for each image, resulting in a different patterns of signal loss (dark areas) due to anisotropic diffusion [7].	42
2-3	Representation of a tensor with an ellipsoid whose principal axes lie on the principal directions of the tensor, \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 . The main axes are proportional to the corresponding eigenvalues, λ_1 , λ_2 , and λ_3 . . .	44
2-4	The fractional anisotropy map in axial and coronal views. The bright regions (higher FA value) represent the regions with dense fibrous tissue as the water diffusivity is highly restricted in the one direction that is parallel to the orientation of fiber tracts in the brain.	46
2-5	Ellipsoidal representation of a tensor with either linear, planar, or spherical symmetry.	47
2-6	A composite representation of a tensor decomposed to its linear, planar, and spherical components.	47
2-7	Example of fiber trajectories extracted using a tractography algorithm on the whole human brain. Tractography seed locations, colors, and transparencies were manually chosen by a neuroanatomist. [63] . . .	48

- 3-1 A schematic representation of a trajectory (\mathbf{r}_{ik}) and a cluster center($\boldsymbol{\mu}_k$). They are equidistance sampled. The \mathbf{r}_{ik}^k represents the trajectory with its point-correspondence to cluster k is determined. 63
- 3-2 (a) Distance map from sample points on a cluster center and (b) the point correspondence label map with the center overlaid. Each region in the label map, displayed by a different color, consists of all of the points in the space that have the minimum distance to a specific point on the cluster center. Therefore, projecting any curve onto this label map determines the point correspondence of each of its samples to the center based on which region that sample is located. 64
- 3-3 Schematic representation of cases where a penalty for shape dissimilarity is required. (a) A trajectory (marked by blue dots) that has a very small distance to the cluster center (marked by red dots), but is not similar in shape to it. (b) A trajectory that has roughly the same distance to either of the cluster centers, but we prefer it to be assigned to the left cluster center that has roughly the same length as the trajectory. 65
- 3-4 Adjustments made to the total distance between the trajectory (blue) and the cluster center (red) calculated by Eq. (3.7). The penalty is equal to the number of missing point times the average distance of the matched points, d_{av} . This is implemented more efficiently as $d_{\text{penalty}} = ((L_k - L_i) + \text{Number of repeated points}) \times d_{av}$, where L_k and L_i are the length of the cluster center and the trajectory, respectively. (a) $d_{\text{penalty}} = 2d_{av}$, (b) $d_{\text{penalty}} = 6d_{av}$, (c) $d_{\text{penalty}} = 0$ 66

4-1	In our approach the probabilistic model is constructed on the distances between the trajectories (solid curves) and the cluster centers (dashed curves), not on the point coordinates. For each pair of trajectory \mathbf{r}_i and cluster center $\boldsymbol{\mu}_k$, a distance d_{ik} is calculated as explained in Chapter 3. As detailed in this chapter, we further assume that the d_{ik} 's follow a gamma distribution parametrized with θ_k , i.e., $d_{ik} \sim \text{Gamma}(\theta_k)$. .	74
4-2	Gamma distribution function with different set of parameters that control the shape of the distribution.	76
4-3	Histogram of the distances of the trajectories from the cluster center and the fitted gamma distribution.	77
4-4	A graphical model of the clustering of fiber trajectories in the absence of an anatomical prior. The observed variables, i.e., shaded node, are the distances from each trajectory to cluster centers. z_i is defined as the label of each trajectory which takes on integer values. \mathbf{g} is the collection of the parameters for the Gamma distributions and \mathbf{w} is the mixing weights.	78
4-5	Schematic representation of procedure used by the EM algorithm to find the local maximum of the log likelihood function, $L(\Theta)$. In the expectation step the algorithm constructs a lower bound $L^t(\Theta)$ based on the current estimate of the parameter, Θ^t , of the optimal solution, $\hat{\Theta}$. In the maximization step, the maximum of the lower bound is calculated to obtain the updated parameter, Θ^{t+1}	80
4-6	Evolution of the Gamma distributions that characterizes the normalized distance metric for the two clusters shown on the left in Figure 5-12. After just a few iterations, the distribution converges to a narrow distribution with small but non-zero mode.	89

4-7	(a) Trajectories of 5 different clusters used for quantitative analysis: splenium (yellow), corticospinal (red), corticobulbar (green), middle cerebellar peduncle (blue), and genu (magenta). Robustness of the EM algorithm with respect to the initial cluster centers is clearly seen. The algorithm was run 5 times with different initial centers (b) to cluster the trajectories in (a). Final cluster centers collected in (c) show little dependence on initial centers.	91
4-8	Clustering of the trajectories (colored based on the local FA value) shown in (a) to Inferior Longitudinal Fasciculus (white), Inferior Fronto-Occipital Fasciculus (red) and Uncinate Fasciculus (blue) clusters shown in (b).	92
4-9	Demonstration of outlier rejection by imposing a threshold on the likelihood function. (a) Unclustered trajectories from superior longitudinal fasciculus. (b) Probabilistic assignment of these trajectories into arcuate fasciculus and the upper portion of SLF. Initial cluster centers are shown in white, while final cluster centers are shown in yellow. (c) Trajectories identified as outliers with a threshold of 0.2 on the likelihood function.	93
4-10	Control of the spread of the bundles with the likelihood threshold. With a threshold of 0.6 more compact trajectories are obtained (b) compared to those obtained with a threshold of 0.2 (a).	94
4-11	About 3000 trajectories clustered into 25 user-initialized bundles. Clusters include different segments of the corpus callosum, tapetum, middle cerebellar peduncle, corticobulbar and corticospinal tracts, and different portions of thalamic radiation.	95
4-12	Cingulum bundles are the two narrow C-shaped structures on the top of the U-shaped corpus collsum, colored in purple and light green. Due to the adjacency to the corpus callosum, many callosal trajectories are included in tractography of these bundles that adversely affect any further analysis on them.	96

4-13	Registration process performed on cingulum trajectories from 8 different cases as a part of the pre-processing from clustering. Sagittal view of the trajectories before (left) and after (right) registration.	97
4-14	Clustering performed on cingulum trajectories from 8 individual cases. For each case the axial view of the trajectories before (top) and after (bottom) clustering is shown. Trajectories are clustered into superior and anterior part of the cingulum, shown in yellow and red, respectively. Note that the clustered bundles in each subject have a similar thickness with a single specified threshold for the population of trajectories.	98
4-15	Processing time as a function of the number of trajectories measured for clustering of the left uncinate fasciculus from 44 cases. Open squares show the whole processing time, where its minimum is limited by the computation of the distance maps. Filled triangles show the processing time minus the time required to compute the distance map, demonstrating a linear dependence on the number of trajectories, i.e., a slope of approximately one in a log-log scale.	99
5-1	The constructed atlas of fiber tracts shown for the obtained bundles in the sagittal (a) and anterior coronal (b) view [58].	103
5-2	Visualization of some of the ROIs outlined by the atlas. These ROIs correspond to the major anatomical fiber tracts.	105
5-3	Demonstration of the registration process for the trajectories of superior cingulum in one of the subjects. Original trajectories (green) and trajectories mapped into the atlas space with either an affine registration (red) or a coealing registration to a common space for all subjects and then by an affine registration into the atlas space (yellow) are shown. The spatial extent of the cingulum specified by the atlas is also shown for comparison.	106

5-4	A Bayesian network that describes the dependencies among different variables. $\boldsymbol{\pi}$: atlas prior, z : trajectory label, \mathbf{d} : distance between trajectory and cluster centers and \mathbf{g} : parameters of the gamma distributions. Shaded nodes represent the observed data.	108
5-5	A comparison of the clustering results without (top) and with (bottom) an atlas. Axial and sagittal views are shown for superior cingulum. Trajectories are colored by their assigned membership probability. The cluster centers at consecutive EM iterations are shown in yellow, and dotted lines represent the initial centers. Without the atlas, and with improper setting of the clustering parameters the cluster centers drift and their extent increases as the algorithm proceeds. Less sensitivity to parameter setting and greater robustness is achieved with the atlas incorporated.	111
5-6	Graphical models for the coin flipping problem. (a) A coin is flipped 10 times and the probability of getting a Head, i.e., $P(H) = \theta$ is inferred from the outcome. (b) The model is extended to add our prior knowledge that most probably the coin is fair. The conjugate prior can be interpreted as some “fictitious data”, i.e., as if we had observed in the past that from $F_H + F_T$ tosses F_H came up Heads. (c) The model is further extended to add a model selection level; whether the coin is fair or not is also inferred from the coin tosses.	112
5-7	A graphical model that shows the inference of the priors and cluster parameters from the data when an atlas prior is available that describes the dependencies between different variables. $\boldsymbol{\pi}$: atlas prior, z : trajectory label, \mathbf{d} : data and \mathbf{g} : parameters of gamma distributions. .	115

- 5-8 Control of the Dirichlet distribution with a weight factor. (a) With only a multiplying factor, γ , the expectation of the distribution is kept constant and only its standard deviation is varied. So, only a limited set of distributions can be used. (b) With a weighted averaging with the unity vector as in (5.13) the distribution takes any desired form from a uniform distribution ($a = 0$) to a highly peaked one ($a \gg 1$). 116
- 5-9 A comparison of the clustering results with two degrees of atlas influence and with an atlas prior that agrees with the membership likelihood inferred from the mixture model: (a) $a = 0$, i.e., no atlas, and (b) $a = 1$. Histograms of the membership probability p_{ik} , are shown for synthetic data colored by the true cluster assignment. The misclassification is indicated by the presence of red samples on the left side of the histograms or blue samples on the right. The ratio of mis-clustered data points decreases from 15% to less than 2% when the atlas is used. Note that the y-axis is in logarithmic scale. 122
- 5-10 A comparison of the clustering results with different degrees of atlas influence, but with an atlas prior that disagrees with the membership likelihood inferred from the mixture model: (a) $a = 0$, i.e., no atlas, (b) $a = 0.25$, (c) $a = 0.5$, and (d) $a = 1$. Histograms of the membership probability, p_{iks} , are shown for synthetic data colored by the true cluster assignment. Misclassification is indicated by the presence of red samples on the left hand side of the histograms. The misclustering ratio increases as the influence of the atlas is increased. Note that the y-axis is in logarithmic scale. 123
- 5-11 Effect of atlas weight, a , on the mis-clustering ratio and when the atlas prior agrees (filled symbols) or disagrees (open symbols) with the membership likelihood inferred from the mixture model. 124

5-12	The influence of atlas weight on the clustering of cingulum trajectories: (a) atlas-specified ROI for cingulum and clustering results with (b) $a = 1$ and (c) $a = 0.5$. When the atlas is imposed strongly some of the trajectories that belong to the cingulum are rejected only because they do not have enough overlap with the atlas ROI.	125
5-13	Demonstration of the performance of the congealing algorithm used to align the FA volumes from 44 subjects to a common space. Axial slices of the mean FA volumes are shown (a) before and (b) after registration. Note the improved sharpness of the plots in (b).	126
5-14	Demonstration of the performance of the affine registration used to map the output of the congealing algorithm to the MNI space. After overlapping the images, misalignment is shown in red (a) before and (b) after applying the affine registration.	128
5-15	Sensitivity of the clustering results to registration errors. Mean square error of the posterior probabilities are plotted as a function in the (a) translation and (b) rotation error. Filled symbols are results of our algorithm, whereas open symbols correspond to an atlas-only clustering method.	129
6-1	(a) Trajectories of 5 different clusters used for quantitative analysis: splenium (yellow), corticospinal (red), corticobulbar (green), middle cerebellar peduncle (blue), and genu (magenta). (b) A model representation of the bundles as the mean trajectory and the isosurfaces corresponding to spatial variation of the clusters.	133
6-2	Curvature of the cluster center along its normalized arc length for fiber bundles shown in Figure 6-1: (a) splenium, (b) genu, (c) middle cerebellar peduncle, (d) corticospinal, and (e) corticobulbar fiber tracts.	134

6-3	FA-colored trajectories from (a), (d) cortico-spinal, (b), (e) cingulum and (c), (f) uncinate fasciculus at 32-wk (up) and 42-wk (down) postmenstrual age. Spatial patterns of the tract development are clearly seen.	136
6-4	Box-plot of the FA variation along the tract arc length for the superior part of the cingulum and at 32-wk (up) and 42-wk (down) postmenstrual age. Only the posterior part shows a significant FA increase. ROI-based analysis fails to detect such spatial dependencies.	137
6-5	(a) Axial and (b) sagittal view of the clustered trajectories from genu and splenium parts of the corpus collasum. Trajectories from all cases are registered into the atlas space, clustered, and colored with the local fractional anisotropy.	139
6-6	The variation of FA along the arc length, from left to right of the brain, for genu (a), upper splenium (b) and lower splenium (c) for healthy (solid lines) and diseased (dashed lines) subjects and the corresponding p-values. Significant differences are observed on the left side of the genu and the lateral portion of the splenium.	140
6-7	Similar to Figure 6-6 but for mean diffusivity along the tract.	141
6-8	(a)The trajectories of three bundles of UF, ILF and IFO in the atlas space are clustered by our algorithm. These bundles are manually shifted in space for better visualization in (b) and are colored by FA values in (c).	143
6-9	Variation of fractional anisotropy along the tract arc length from inferior to anterior for (a) left and (b) right uncinate fasciculus in healthy (solid lines) and diseased (dashed lines) subjects and the corresponding p-values (c) and (d).	144
6-10	Variation of (a) parallel diffusivity and (b) perpendicular diffusivity along the tract arc length from inferior to anterior for (a) left and (b) right uncinate fasciculus in healthy (solid lines) and diseased (dashed lines) subjects and the corresponding box plots (c) and (d).	144

6-11	Comparison of the fractional anisotropy along the normalized arc length of the left (solid lines) and right (dashed lines) UF and for the (a) control and (b) schizophrenia cases. The corresponding p-values in (c) and (d) demonstrate statistically significant left-greater-than-right asymmetry in the inferior part and right-greater-than-left asymmetry in the superior part of the UF in both groups.	145
6-12	Comparison of the fractional anisotropy for (a) left and (b) right ILF for control (solid lines) and schizophrenia (dashed lines) groups, along with the corresponding p-values (c) and (d). No significant change in FA is observed for these bundles.	146
6-13	Comparison of the (a) perpendicular diffusivity in the left ILF and (b) parallel diffusivity in the right ILF for the healthy (solid lines) and schizophrenia (dashed lines) groups, along with the corresponding p-values (c), (d).	147
6-14	Comparison of the fractional anisotropy for the (a) left and (b) right IFO and for the healthy (solid lines) and schizophrenic (dashed lines) groups, plotted against the normalized arc length of the bundle. The corresponding p-values are shown in (c) and (d).	148
6-15	Comparison of the (a) parallel and (b) perpendicular diffusivity for the right IFO in healthy (solid lines) and schizophrenia (dashed lines) groups. Significant increase in the perpendicular diffusivity is observed for a large portion of this bundle.	148
6-16	Visualization of samples of cluster centers for (a) left and (b)right UF, ILF and IFO bundles. The radius of each sample is proportional to the significance of the difference between the FA value of the schizophrenics and controls at that point. The increase and decrease in FA is highlighted in red and blue respectively.	149
6-17	Clustered trajectories of (a) left and (b) right cingulum from 44 subjects are shown in the atlas space. (c) shows these bundles from the axial view which each voxel is colored based on its FA value.	151

6-18	Comparison of the fractional anisotropy along the arc length of the superior part of the (a) left and (b) right cingulum in control (solid blue) and schizophrenia (dashed red) cases and (c), (d) the corresponding p-values. Increased FA is observed in the middle of left cingulum, while decreased FA is seen in the anterior part of the right cingulum.	152
6-19	Comparison of the (a) parallel and (b) perpendicular diffusivity along the arc length of the superior part of the right cingulum in control (solid blue) and schizophrenia (dashed red) cases and (c), (d) the corresponding p-values. The anterior part of the bundle shows increased diffusivity in all directions. Hence the FA analysis in Figure 6-18(b) does not completely reflect the changes in the right cingulum.	152
6-20	Comparison of the fractional anisotropy along the normalized arc length of the superior part of the left (solid lines) and right (dashed lines) cingulum and for the (a) control and (b) schizophrenia cases. The corresponding p-values in (c) and (d) demonstrates statistically significant left-greater-than-right asymmetry in both groups. A more important observation is that the asymmetry is more pronounced in the schizophrenic cases.	153
6-21	Three-dimensional visualization of the significance of the group analysis performed for the FA values between the control and schizophrenia subjects.	153
6-22	Age-dependence of the fractional anisotropy in mid-sagittal point of the lower splenium for healthy (open squares) and schizophrenic (closed squares) subjects.	155
A-1	Different situations where nodes X_1 and X_3 are either dependent or independent, based on their connection to X_2	161

List of Tables

2.1	Categorizing DT-MRI analysis studies based on the quantity being measured. In each case the required tools and the approaches taken to develop these tools are listed.	52
2.2	Categorization of DT-MRI analysis work based on where the quantitative parameters are measured and compared. In each case the required tools specifying the location and the approaches proposed to address these requirements are listed.	53
3.1	Comparison of the unidirectional and bidirectional Hausdorff distance with the distance measure introduced in this work for typical trajectory-center pairs. Blue points represent samples of the trajectory, while red points are samples of the cluster center.	68

Chapter 1

Introduction

The human brain is composed of processing units which are connected by millions of neurons. The transfer of information in a neuron is done through its wire-like part, called the axon, which carries the electric signals away from the neuron's body and enables communication with other neurons through synaptic processes. Bundles of these axons, which are called fiber tracts, play an important role in the proper functioning of the brain. Many neural diseases are hypothesised to be associated with disruption and damage to these fiber tracts in the white matter. Understanding the pattern of connectivity of fiber tracts, their properties in healthy subjects, and how they are affected in each diseased population is of great interest to neuroscientists, neurosurgeons, and the medical community in general. *In vivo* study of the white matter fiber tracts was not possible by traditional imaging modalities, as they were not able to capture the network of microscopic cellular wire-like structures. It is only for the last decade that diffusion-weighted magnetic resonance imaging (DW-MRI), and its variant Diffusion Tensor MRI (DT-MRI), have enabled us to reconstruct and study these connecting paths in the brain.

Unlike conventional magnetic resonance imaging (MRI), which measures quantities related to tissue composition, DW-MRI measures the diffusivity of water. Given that the human brain is an anisotropic medium, in which the water movement is restricted in the direction perpendicular to the bundles of axons, diffusion MRI is widely used to study the orientation and integrity of these bundles *in vivo*.

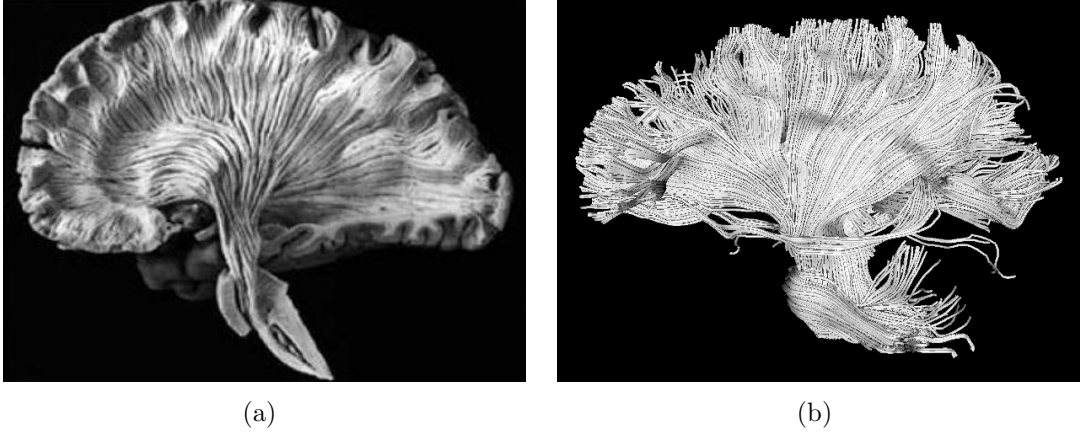


Figure 1-1: (a) A sagittal dissection of postmortem brain that shows white matter fiber tracts (From Virtual Hospital [98]). (b) An example of fiber trajectories reconstructed from diffusion tensor images.

Figure 1-1(a) shows a sagittal view of a postmortem brain in which the fiber tracts can be easily seen. Figure 1-1(b) depicts a set of three-dimensional curves, produced from diffusion tensor MRI by a tractography scheme. Tractography methods trace the trajectories of fiber tracts by following the direction of water diffusion and provide a powerful tool to visualize the extent and orientation of the fiber tracts and to study brain connectivity. The extracted trajectories can also be used to study brain integrity, as the local diffusion parameters and the shape of the trajectories are affected by brain diseases and brain development and aging.

Due to the importance of this relatively new imaging modality and the inadequacy of methods developed for conventional MRI to analyze diffusion MRI, there has been an increasing number of methods proposed in recent years for processing diffusion MRI data. This thesis presents a novel framework to perform quantitative analysis along fiber trajectories and to include anatomical information in such an analysis at different levels of abstraction.

Diffusion MRI has gained tremendous popularity in imaging of the white matter of the brain. The extension of available technology will aid in the diagnosis and subsequent treatment of disorders of the central nervous system and is likely to have a major impact on assessment of white matter pathologies (e.g., multiple sclerosis), quantification of abnormal white matter development, detection of stroke and trauma,

trauma at brain swelling, diffuse axonal injury, and spinal trauma, as well as a large variety of brain tumors. In addition to direct clinical impact, the extension of available diffusion MRI technology will also contribute to basic neurosciences, improving our understanding of physiological white matter development, mature brain connectivity, and aging.

1.1 Quantitative Analysis of Diffusion MRI and Its Challenges

In the literature, to date, three major approaches have been pursued in analyzing diffusion MRI data:

- Region of interest (ROI)-based methods,
- Voxel-based methods, and
- Tract-oriented methods.

ROI-based methods are by far the most popular in clinical studies. Parameters, such as fractional anisotropy (FA), are measured in manually or semi-automatically defined regions of interest and averaged over groups of healthy and diseased populations. The main advantage of ROI-based methods is that they provide an easy route to examine hypotheses related to the role of a particular tract associated with a specific brain disorder. However, such ROI-based methods are not time-efficient as they often require user interaction to specify the ROIs and their accuracy is limited by the reliability of specifying the ROIs. Most studies use FA or structural images to specify the ROIs. This would be very difficult and prone to significant errors for structures that do not fall on an image plane or are located in the vicinity of other fiber tracts. Even for structures where ROI placement is relatively easy, it has been shown that ROI size, shape, number, and location not only affect the measured quantities, but also influence the significance of the group analysis [47]. A review of diffusion MRI studies in schizophrenia, for example, shows considerable variation in the results.

While this might be partly due to sample size and group matching, variations in ROI localization cannot be ruled out [47].

In voxel-based quantitative analysis, the data sets are registered into a common coordinate system and then averaged and compared voxel-by-voxel. This is algorithmically simple, but it is critical to obtain a good image alignment [47], which otherwise directly translates to error in the quantitative analysis. The simplest approach is to register scalar fields, such as fractional anisotropy, where methods already developed to register conventional MR images can be used. However, a scalar measure does not employ all of the information contained in the diffusion data and thus would not provide the most accurate analysis. To use the full diffusion data in the registration, non-linear warping of the tensor field [77] is required, which in turn needs re-orientation of the tensors [2, 67]. The main advantage of voxel-based methods over ROI-based approaches is that they are, in principle, user-independent and are suited for whole brain analysis. The latter is particularly advantageous when no hypothesis regarding the location of brain changes is available *a priori*. However, apart from the registration errors and partial volume effects, especially in the regions with high anisotropy contrast, associating the observed differences to specific tracts is difficult [47]. The data are often smoothed to ensure the validity of the statistical analysis. This results in an inherently low resolution map of the differences and voxels exhibiting significant group differences do not necessarily lie within an anatomical tract. Reference to an anatomical atlas is hindered by the low resolution of the obtained difference map and by the limited resolution of the atlas itself [47].

An alternative approach is to compute the parameters of interest along fiber trajectories. Even though tractography algorithms still suffer from several shortcomings (e.g., cannot deal properly with fiber crossings), their clinical application is increasing. Visualizing the fiber tracts as a collection of trajectories has attracted neuroscientists as it provides an easy way to display the extent and direction of the underlying tracts. Also, noise and image imperfections that are present in DT images are translated to outlier trajectories that can be identified visually. We believe tract-oriented quantitative analysis offers advantages over ROI-based methods since it is able to reveal the

local variations of the fiber integrity, which are lost when the quantitative parameters are averaged over the entire fiber tract in ROI-based methods. Tract-oriented analysis is valuable especially if performed on a group of trajectories that correspond to a single anatomical fiber tract. In doing so, first an algorithm is needed to group fiber trajectories into clusters that correspond to anatomical bundles. The principal benefit that clustering offers is that when comparing two subjects or populations, one compares the same tracts. Observed differences are then due to differences in the properties of the specific tracts rather than differences in the overall anatomy/shape of the individual brains.

Despite the benefits and potential applications of tract-oriented analysis, many challenges of this approach have not yet been addressed.

The input data for a tract-oriented quantitative analysis is a group of trajectories generated by a tractography algorithm. Regardless of the nature of the tractography method, whether deterministic or probabilistic, the output trajectories often have discontinuities due to the presence of noise and image imperfections. Even with an ideal tractography algorithm and elaborate preprocessing of the data to remove noise and imperfections, brain abnormalities such as lesions may cause discontinuities in the trajectories near the brain abnormality. As a result, the clustering algorithm should be able to deal with variable-length input data.

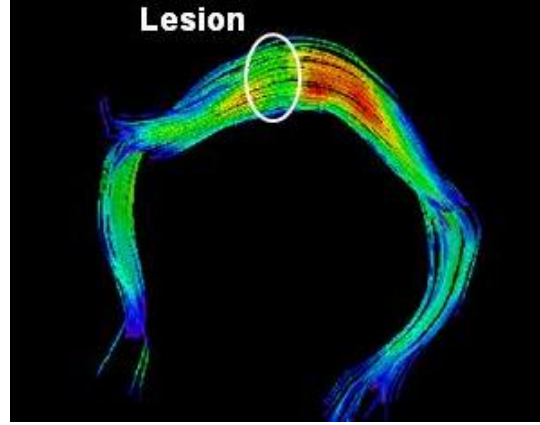
Another issue in clustering is defining the similarity between the trajectories. In general, the similarity between three-dimensional curves (trajectories) is not uniquely defined and depends on the application. In our case, we require the similarity measure to use both spatial and shape information from the whole trajectory. The common approach in defining the similarity is to first find the correspondence between the points on each curve. Since the number of trajectories is typically quite large, a computationally efficient implementation is required. Three-dimensional curve matching requires pairwise comparison of the trajectories (and their sub-segments) or comparison between the trajectories and the cluster centers (and their sub-segments) and thus it is not time-efficient. An alternative to uniform sampling of the trajectories to establish the point correspondence is to use landmarks that contain most of the

information about the shape and location of the trajectory. Such landmarks can be defined, for example, through a minimum description length approach [26] or as the local extrema of the curvature [29]. When dealing with trajectories that have similar shape, such landmarks can be taken as the corresponding points. However, adaptation of such approaches to define the point correspondence in our application, where some features might be missing from the individual curves, requires extensions of those algorithms.

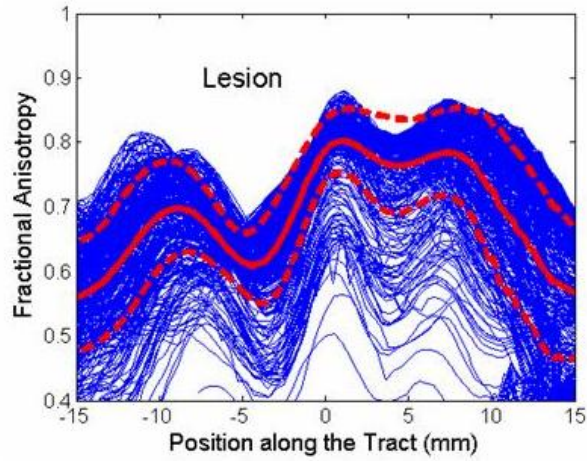
Regardless of the defined similarity measure, an unsupervised algorithm is not guaranteed to produce the clusters of interest for a given application. The user only has control over the number of clusters and the algorithm could easily over-cluster or under-cluster the input data. We believe that a supervised clustering algorithm that benefits from anatomical information is needed. Once anatomical data is used in the clustering step, the correspondence between clusters across different subjects, needed for population studies, is automatically known. In this thesis we address the problem of how to incorporate anatomical information in the clustering algorithm.

Once the trajectories are grouped into clusters, the aim of the quantitative analysis is to provide statistics of the desired quantitative parameters, such as different diffusion anisotropy measures, for the trajectories that belong to a given cluster. A tract-oriented quantitative analysis reports the average (and higher-order moments) of the local parameters along the tracts. It thus requires that the statistical calculation is performed over trajectory points that correspond to each other. However, since the trajectories that belong to a given cluster do not necessarily have the same length (or their start points do not correspond to each other), point correspondence between the trajectories first needs to be determined. Unlike the point correspondence required in the clustering step, this is mandatory and should be calculated carefully.

The framework presented in this thesis is based on trajectories estimated from diffusion data by a tractography algorithm, along with the raw diffusion measurements of each image. The algorithm can also use some anatomical information in the form of a set of representative trajectories and/or soft assignment of the input trajec-



(a)



(b)

Figure 1-2: An example of tract-oriented quantitative analysis performed on one part of the corpus callosum of a subject diagnosed with multiple sclerosis. (a) A group of the fiber trajectories are colored based on the local fractional anisotropy (FA). (b) FA variation along the trajectories (blue) along with mean (solid red), and mean plus/minus standard deviation (dashed red) of the FA. The tract-oriented analysis identifies the FA reduction in the vicinity of the lesion without prior knowledge of where the lesion is located.

ries to anatomical fiber bundles. The output is the probabilistic assignment of the input trajectories to anatomically-known bundles of fiber tracts and the statistical measurements of the quantitative parameters along the trajectories. The analysis can be performed either on a single subject or over a group of subjects. Figure 1-2 shows an example of such analysis performed on one part of the corpus callosum in a subject diagnosed with multiple sclerosis (MS). Part (a) of the figure shows a group of fiber trajectories colored based on a local scalar measure of diffusivity called fractional anisotropy, while part (b) shows the FA variation along the arc length of the trajectories. By such a tract-oriented analysis, we are able to identify a significant drop in fractional anisotropy in the vicinity of the lesion, without knowing where the lesion is located *a priori*.

Figure 1-3 shows the analysis performed on the posterior part of the corpus callosum, called the splenium, for a group of subjects. In (a) the trajectories extracted by a tractography algorithm are shown, (b) and (c) illustrate the results of clustering those trajectories into the upper and lower parts of the splenium, in the sagittal and axial view respectively, (d) shows the variation of FA along the lower bundle, while the corresponding p-values (a measure of statistical significance) are shown in part (e). A clustering procedure is often required to be performed on the output of the tractography algorithms to group similar trajectories together, and to remove outliers.

1.2 Objective and Contributions of The Thesis

We believe that the medical community has not yet fully benefited from what diffusion imaging offers and this is in part due to the fact that well-established robust quantitative methods for analysis of diffusion MRI data are largely missing. This thesis proposes a framework to enable the quantitative analysis and in particular to answer the following questions:

- How to obtain the point correspondence between the trajectories efficiently?

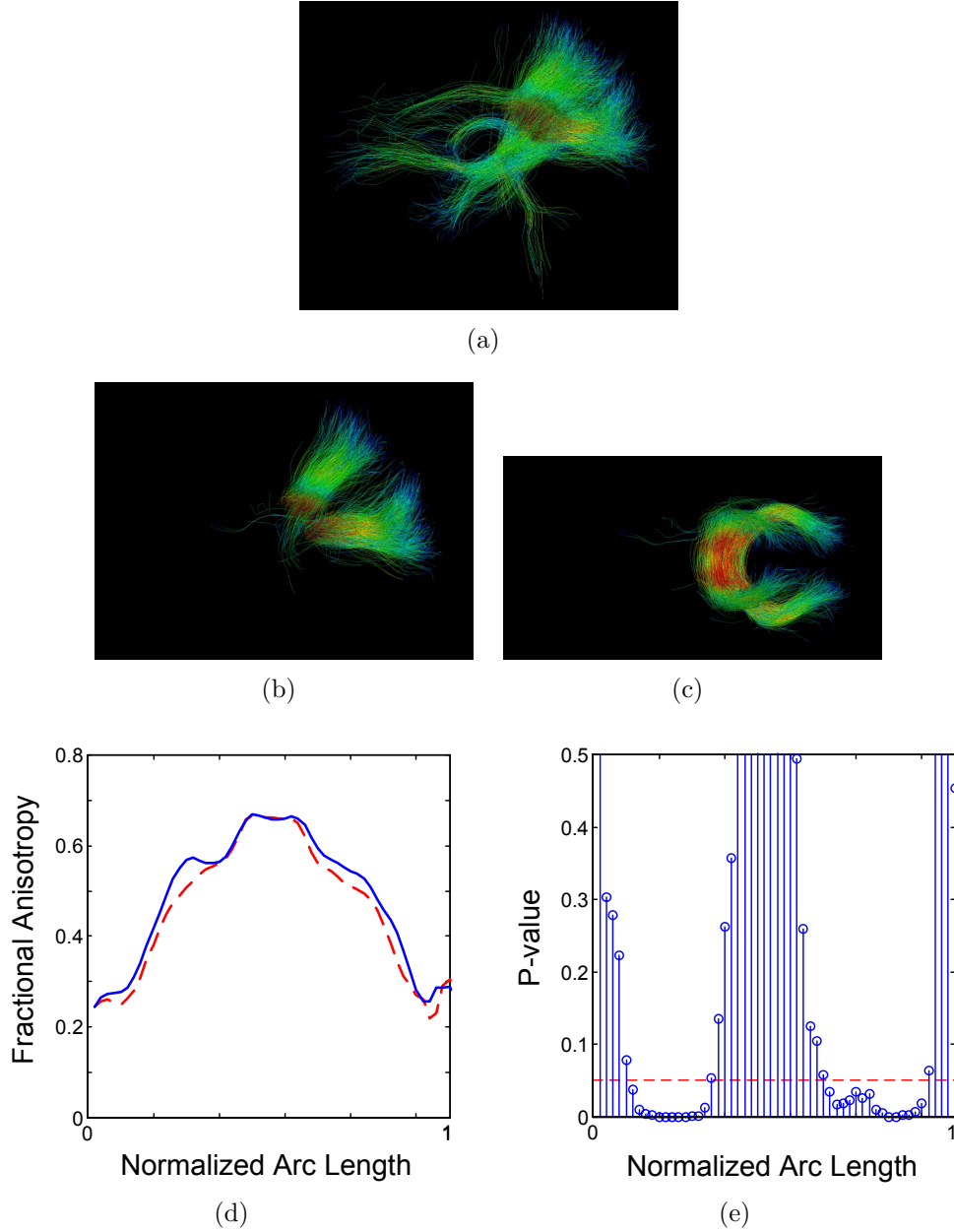


Figure 1-3: An example of tract-oriented quantitative analysis on a population and for the splenium parts of the corpus callosum. (a) Sagittal view of the trajectories from all cases, registered into the atlas space, and colored with the local fractional anisotropy. (b) Sagittal view of the trajectories in (a) that are clustered into the upper and lower parts of the splenium. (c) Axial view of the trajectories shown in (b). (d) Average fractional anisotropy along the tract arc length for the normal (blue) and schizophrenia (red) cases, where a drop in FA is seen for diseased cases. (e) A plot of the corresponding p-values which indicates that a significant drop in the FA is only observed in the mid-portion of each side of the splenium.

- How to define the similarity between 3-D trajectories?
- How to incorporate anatomical information in clustering of the trajectories?
- How to perform tract-oriented analysis over a population?

The answer to the last question is the ultimate goal of this thesis. It is worth mentioning again that being able to perform such analysis is valuable since it enables us to study the local changes of quantitative parameters along the fibers. This is especially interesting for studying temporal changes of fiber tracts during brain development and it also opens new possibilities to compare normal and pathological subjects.

As an answer to the first question mentioned above, we propose a novel approach for calculating the point correspondence between the trajectories by building a distance map and a Voronoi diagram on the same space. This provides a computationally efficient alternative to curve matching algorithms when dealing with a large number of curves (feature vectors) when performing population analysis. This enables us to use the point correspondence in the clustering step. Since the number of trajectories is quite high, other approaches sacrifice the accuracy of determining point correspondence either by using a simple similarity measure or by using information from a limited number of trajectory points.

Having the knowledge of point correspondences, we define a similarity measure in which spatial information is used explicitly while shape similarity is included implicitly through penalty terms for missed and repeated point matches.

Theoretical models are also developed to include anatomical information from an atlas of fiber tracts. We describe two levels for including the anatomical prior. In the first approach, the priors are fixed and are given by an anatomical atlas. The second approach uses a Dirichlet distribution to control the influence of the atlas. To the best of our knowledge these are the first implementations in which anatomical priors are used in a mathematically principled framework. By using an atlas, the correspondence between clusters in different subjects is automatically known.

To ensure robustness, quantitative analysis is performed probabilistically using the membership weights (probabilities) assigned to each trajectory in the clustering step. In the mixture model clustering implemented in this work, we develop the idea of building the mixture model based on the *distance* between the trajectories and the cluster centers to deal with the variable length of the trajectories. The output of clustering is the probabilistic assignment of trajectories to anatomically-known bundles of fiber tracts and the point correspondences between trajectories to the mean trajectory of each cluster. This information is used to compute proper statistics on any attribute feature vector, such as scalar diffusion measures, e.g., fractional anisotropy and mean diffusivity, defined along the trajectories. In population studies, the trajectories are first transformed to a common space, clustered, and the quantitative parameters of interest are reported along the tracts.

As will be shown in Chapter 6, the proposed tract-oriented quantitative analysis addresses the following clinical questions:

- To what extent and where on a white matter fiber tract the diffusion parameters change between two time points during the brain development?
- Is there diffusion asymmetry between the two brain hemispheres and how are such asymmetries affected by diseases?
- Is there any localized group difference in a given fiber tract between normal and diseased populations?

As a byproduct of the proposed clustering method, a spatial model of the fiber bundles represented by the mean trajectory and its spatial variation is also obtained. This is shown in Figure 1-4 in which the abstract models of five fiber bundles are visualized by their spatial mean and iso-surfaces corresponding to the mean plus three standard deviations (3σ) of the 3-D coordinates calculated along the cluster center. Such an abstract spatial model for fiber bundles could be used for neurosurgery applications. It enables one to easily visualize the extent of the fiber tracts adjacent to the brain lesions to help minimize the damage to the bundles when removing

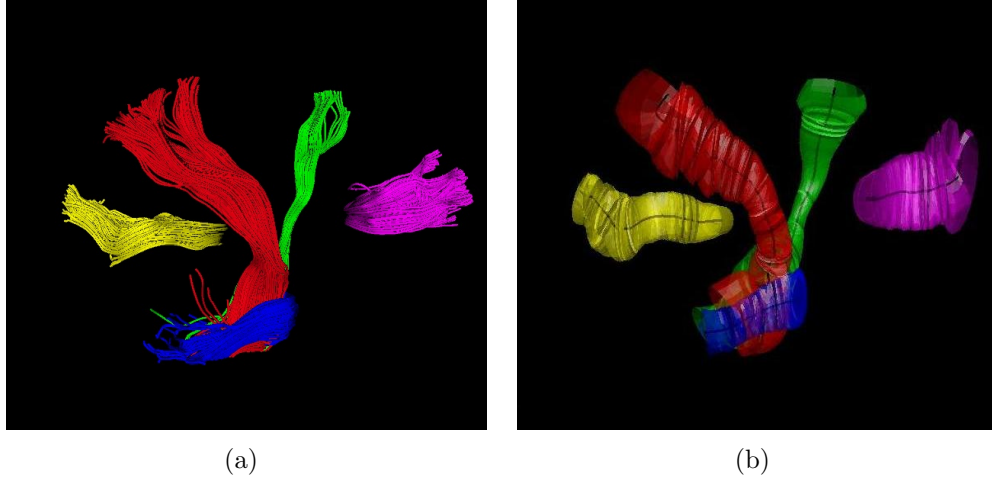


Figure 1-4: (a) Trajectories of 5 different clusters: splenium (yellow), corticospinal (red), corticobulbar (green), middle cerebellar peduncle (blue), and genu (magenta). (b) A model representation of the bundles as the mean trajectory and the isosurfaces corresponding to spatial variation of the clusters.

the lesion. Conventional DT-MRI-based approaches usually visualize a map of the fractional anisotropy [50] or bundle segmentations based on the FA map. However, such approaches are not accurate at fiber crossings where FA is low. An alternative approach is to visualize the fiber trajectories [44], but rendering such a large dataset can be confusing and not be efficient in real-time applications.

1.3 Organization of the Thesis

We provide the background on DT-MRI quantitative analysis in the next chapter. Hypotheses that govern the relationship between the apparent water diffusivity measured by DT-MRI and the axonal density, orientation, and integrity are presented. Then quantitative measures of diffusion are introduced, followed by a review of previous work on DT-MRI quantitative analysis.

In Chapter 3, we introduce a new method for efficiently calculating similarity between 3D curves. Limitations of the methods commonly used by the DT-MRI community for calculating the similarity between fiber trajectories are discussed and then the proposed method is introduced.

Chapter 4 presents the probabilistic clustering method developed in this work

to group white matter fiber trajectories into anatomically meaningful bundles. An expectation-maximization algorithm is used to cluster the trajectories in a mixture-model context. One contribution of our method is to perform the mixture modeling on the *distances* between the trajectories and cluster centers as opposed to the coordinates of trajectories. We demonstrate how this technique performs well on variable-length feature vectors. We also demonstrate that unlike the Gaussian distribution which is the common choice in density estimation, a gamma mixture model accurately represents the nature of the distance of the 3-D trajectories from the cluster centers.

Chapter 5 extends the method presented in the preceding chapter to incorporate anatomical information in clustering of the fiber trajectories, while the influence of the prior on the clustering can be well-controlled by the user. To the best of our knowledge, these are the first implementations in which an anatomical prior is used in a mathematically principled framework.

Some applications and new findings of the proposed algorithms for clinical studies, such as a group analysis on schizophrenia are introduced in Chapter 6.

Chapter 7 concludes the thesis by summarizing major contributions and findings. Suggestions for future work are also discussed in this chapter.

Chapter 2

Background on Diffusion MRI

Quantitative Analysis

Diffusion weighted MRI (DW-MRI) is an imaging modality that measures the local diffusivity of water *in vivo*. Since water diffusion is restricted in the direction perpendicular to neuronal axons, the diffusion tensor, as will be described below, provides valuable information about the local density and orientation of the axons, axonal damages, and myelin damages. Hence DW-MRI is a powerful tool for studying brain organization, development, and diseases. This chapter provides background material on quantitative analysis of diffusion images. It first presents the correlation between the apparent diffusivity of water molecules measured by diffusion tensor MRI (DT-MRI), a variant of DW-MRI, and the axonal organization of the brain white matter. Then basics of DT-MRI through diffusion weighted MRI acquisition are presented. Since three-dimensional tensor data are not easy to visualize, several scalar measures have been introduced by the DT-MRI community and are reviewed in this chapter. We also describe *tractography*, as a tool that to date has mostly been used in visualizing diffusion MRI data. We discuss three major algorithms for tractography to provide some background. However, it is worth mentioning that the algorithms we introduce in this thesis for DT-MRI quantitative analysis take the output of the tractography as their input and hence are independent of the details of the tractography. We finish this chapter with a review of related work on quantitative analysis of white

matter fiber tracts.

2.1 Diffusion Anisotropy in White Matter

When there is a concentration inhomogeneity of a substance, molecules of that substance move from regions of higher concentration to other areas through random (Brownian) motion. This thermodynamic process is called diffusion and in a macroscopic scale is governed by Fick's laws [34], which read

$$J = -D\nabla\phi \quad (2.1)$$

$$\frac{\partial\phi}{\partial t} = \nabla \cdot (D\nabla\phi) \quad (2.2)$$

Fick's first law relates the diffusion flux, J , to the gradient (∇) of the concentration ϕ through the diffusion coefficient or diffusivity D , when the concentration within the diffusion volume does not change, i.e., in steady-state diffusion. In the non-steady state condition, the second law can be used to get the rate of the concentration change by taking the divergence ($\nabla \cdot$) of the diffusion flux. When the medium is isotropic, i.e., diffusivity is independent of the direction, the probability to find a molecule at a distance X from its initial location is given by a Gaussian distribution:

$$P(X, t) = \frac{1}{\sqrt{4\pi t D}} \exp\left(-\frac{X^2}{4Dt}\right), \quad (2.3)$$

where t is the diffusion time.

The macroscopic diffusion process described by Fick's laws refers to the diffusion of one substance in another. However, they can also be used to describe the random motion of the molecules when there is only one type of substance. In this case, the process is called self-diffusion and is described in terms of the probability of finding a particle in a certain position and at a particular time. Suppose that a molecule is initially at \mathbf{r}_0 and we are interested in the probability of finding the molecule at \mathbf{r} at a time t , $P(\mathbf{r}|\mathbf{r}_0, t)$. In the case of free self-diffusion, the probability is independent of the initial position and can be applied to all molecules. The probability can thus be

stated in terms of the relative displacement, $\mathbf{R} = \mathbf{r} - \mathbf{r}_0$, and is described by Fick's laws:

$$\frac{\partial P}{\partial t} = \nabla \cdot (D \nabla P(\mathbf{R}, t)), \quad (2.4)$$

where D is called the self-diffusion coefficient.

What makes diffusion imaging of the brain interesting is that the diffusivity of water measured in the brain depends not only on the position, but also on the direction¹.

Diffusion anisotropy is mostly due to the fact that in some regions of the white matter, axons are highly aligned and form fiber bundles. Figure 2-1 shows a schematic view of a myelinated axon. Myelin, the axonal membrane, microtubules, and neurofilaments are all oriented in the direction of the axon and could restrict water diffusion. Hence, the diffusion coefficient perpendicular to the axon, $D(\perp)$, will be smaller than the one parallel to the axon, $D(\parallel)$. It should be noted that millions of axons fit into a voxel of the diffusion tensor images. Hence, diffusion imaging measures the local diffusivity averaged over the volume of a voxel. In regions where axons are aligned and form fiber bundles, a highly anisotropic diffusion coefficient is measured, while in regions where axons are oriented in random directions or two or more fiber bundles cross each other, the diffusion coefficient is less anisotropic.

Although many clinical studies correlate changes in diffusion anisotropy solely to changes in myelination, several experiments indicate that in some cases this is not a correct hypothesis. For example, studies of brain development demonstrate that diffusion anisotropy is observed well before myelination [73]. Moreover, it has been shown that axonal damage results in reduced diffusivity parallel to the axon [12]. The break-down of the longitudinal axonal structure and the build-up of cellular debris is responsible for reduced parallel diffusion. Hence, it is postulated that only when reduced $D(\perp)$ is accompanied by no change in $D(\parallel)$, it can be interpreted as an indication of myelin damage [83].

¹Strictly speaking, what is measured by diffusion MRI is the apparent diffusion coefficient, ADC, which is a function of the diffusion time, geometry of the local volume, and might also be affected by non-diffusive processes including tissue perfusion and fluid flow.

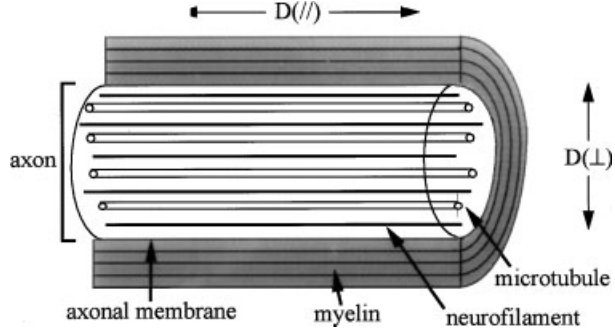


Figure 2-1: A schematic view of a myelinated axon [12]. Myelin, the axonal membrane, microtubules, and neurofilaments are all oriented in the direction of the axon and could restrict water diffusion. Hence, the apparent diffusion coefficient perpendicular to the axon, $D(\perp)$, will be smaller than the one parallel to the axon, $D(\parallel)$.

2.2 Diffusion Weighted MRI

Diffusion Weighted MRI is an imaging modality which utilizes the measurement of Brownian motion of molecules. In diffusion-weighted images, instead of a homogenous magnetic field, the field is varied linearly by a pulsed field gradient. Since precession is proportional to the magnet strength, the protons begin to precess at different rates, resulting in dispersion of the phase and signal loss. Another gradient pulse is applied in the same direction but with opposite magnitude to refocus or rephase the spins. The refocusing will not be perfect for protons that have moved during the time interval between the pulses, and the signal measured by the MRI machine is reduced. This reduction in signal due to the application of the pulse gradient can be related to the amount of diffusion that is occurring through the Stejskal-Tanner equation [85].

With a gradient vector \mathbf{g} , the image intensity, S , is given by [7]:

$$S = S_0 e^{-bD_g}, \quad (2.5)$$

where S_0 is the image intensity in the absence of a diffusion-sensitizing gradient, D_g is the diffusivity in the direction of \mathbf{g} , and b is the diffusion weighting factor in s/mm^2 and depends on the detail of the imaging sequence and the magnitude of the gradient

field:

$$b = \int_0^\Delta \gamma^2 |\mathbf{g}(t)|^2 t^2 dt, \quad (2.6)$$

where γ is the proton gyromagnetic ratio (42 MHz/Tesla), $|\mathbf{g}|$ is the magnitude of the diffusion sensitizing gradient, and Δ is the time between the gradient pulses. It follows that [97]:

$$b = \gamma^2 \delta^2 \left(\Delta - \frac{\delta}{3} \right) |\mathbf{g}|^2, \quad (2.7)$$

where δ is the duration of the gradient pulses. In practice, S_0 is usually acquired with small b values, such as 5 s/mm², while the diffusion weighted signal, S , is measured with a b -value ranging from 500 to 1500 s/mm². Simple calculations [25] show that in the case of isotropic diffusion the best accuracy is obtained if the difference between the b -factors of the two acquisition is roughly $1/D$. For brain DW-MRI this translates to a b -factor of 1000–1500 s/mm². If more than two acquisitions are to be used, it is shown that it is better to repeat the acquisitions at two b -factors instead of using a range of different b -factors [25].

Figure 2-2 shows two baseline images, obtained by applying no gradient, and the diffusion weighted images of a single axial slice.

2.3 Diffusion Tensor MRI

In an anisotropic media, such as brain white matter, the direction dependence of the diffusivity can be, in most cases, described by a tensor². The assumption that the direction-dependence of the diffusivity can be described by a tensor is equivalent to the assumption that the probability of the molecular displacement is given by a multivariate Gaussian as described in [9]:

$$P(\mathbf{R}|t) = \frac{1}{\sqrt{4\pi t^3 |\mathbf{D}|}} \exp \left(- \frac{\mathbf{R}^T \mathbf{D}^{-1} \mathbf{R}}{4t} \right) \quad (2.8)$$

²It has been shown that a single tensor does not describe the orientation dependence of the diffusivity in some cases, e.g., at fiber crossings. Methods such as high angular resolution diffusion imaging (HARDI) [90], diffusion spectrum imaging, or Q-ball imaging [89] have been proposed to deal with such situations.

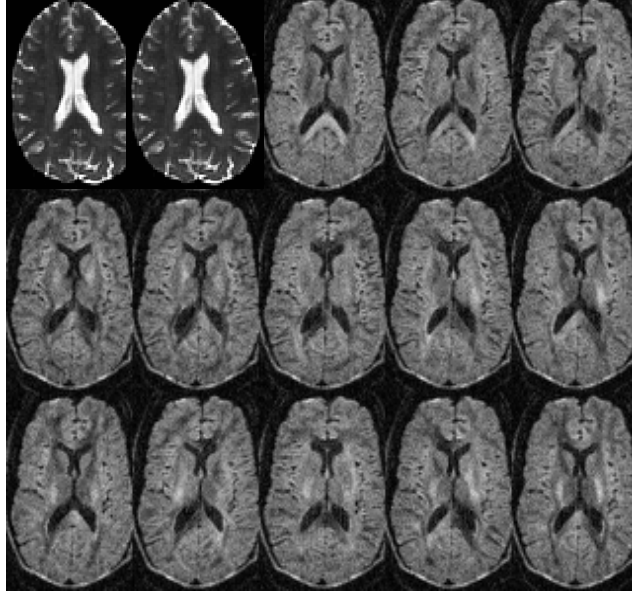


Figure 2-2: Samples of diffusion weighted images acquired from human brain. The first two images are baseline images taken with small b -factor, while the rest are diffusion weighted images. The gradient direction is different for each image, resulting in a different patterns of signal loss (dark areas) due to anisotropic diffusion [7].

where \mathbf{D} is the diffusivity tensor and \mathbf{R} is the displacement vector. In a three-dimensional space, the diffusivity tensor is a positive definite 3×3 tensor and contains 6 independent components:

$$\mathbf{D} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix}. \quad (2.9)$$

So, at least 6 independent gradient directions are needed in DW-MRI to be able to reconstruct the diffusion tensor from the Stejskal-Tanner equation:

$$\log S_k = \log S_0 - b \hat{\mathbf{g}}_k^T \mathbf{D} \hat{\mathbf{g}}_k \quad (2.10)$$

where $\hat{\mathbf{g}}_k$ is the unit vector in the direction of the k th gradient vector and S_k is the corresponding DW image. When there are only 6 gradient directions, the diffusion tensor at each voxel can be obtained by solving the above equation set analytically

[97]. To enhance image quality, today’s image acquisitions often use a larger number of gradient and baseline images, and an optimization method, such as least mean square error, is then used to reconstruct the diffusion tensor [9]:

$$\mathbf{D} = \arg \min_{\mathbf{D}} \sum_k \left(\hat{\mathbf{g}}_k^T \mathbf{D} \hat{\mathbf{g}}_k + \log S_k - \log S_0 \right)^2. \quad (2.11)$$

2.4 Tensor Visualization and Scalar Measures of Diffusion

A tensor can be visualized as an ellipsoid with main axes aligned with the principal or self directions of the tensor as shown in Figure 2-3. Mathematically, this is justified by the fact that being a positive definite tensor, \mathbf{D} can be diagonalized to produce

$$\mathbf{D} = \mathbf{E}^{-1} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \mathbf{E}, \quad (2.12)$$

where λ_1 , λ_2 , and λ_3 are the eigenvalues of the tensor and \mathbf{E} is the matrix whose columns are the corresponding eigenvectors, \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 .

An ellipsoid, in fact, represents the distance in space covered by molecules in a given diffusion time T_d . This follows from the fact that the solution to the Fick’s equation is a Gaussian.

$$\frac{x'^2}{2\lambda_1 T_d} + \frac{y'^2}{2\lambda_2 T_d} + \frac{z'^2}{2\lambda_3 T_d} = 1, \quad (2.13)$$

where x' , y' , and z' are the dimensions in the frame of the principal directions of the tensor (its eigenvectors), and λ ’s are the corresponding eigenvalues.

Such ellipsoids can be visualized at each voxel to represent the direction and magnitude of the diffusivity [97]. However, since such representations are complex, often scalar measures of diffusivity and diffusion anisotropy are calculated and visualized. The most common scalar measures are: mean diffusivity, which characterizes the

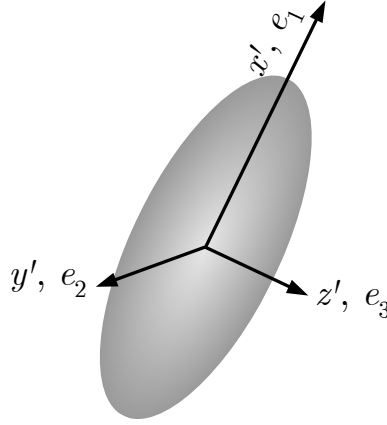


Figure 2-3: Representation of a tensor with an ellipsoid whose principal axes lie on the principal directions of the tensor, \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 . The main axes are proportional to the corresponding eigenvalues, λ_1 , λ_2 , and λ_3 .

overall mean-squared displacement of the molecules or the ellipsoid size; fractional anisotropy, which shows the degree of orientation-dependence of the diffusivity; and color-coding of the orientation. Such scalar measures allow the tensor data to be visualized as grayscale or color images in the same manner used to visualize conventional MRI data.

2.4.1 Mean Diffusivity

Mean diffusivity (MD) is a measure of the average diffusivity of the water in a voxel and is invariant to orientation. Among several combinations of the tensor elements, the trace of the diffusion tensor is such an invariant:

$$Tr(\mathbf{D}) = D_{xx} + D_{yy} + D_{zz}. \quad (2.14)$$

The mean diffusivity is defined as $MD = Tr(\mathbf{D})/3$ and can be re-written as:

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (2.15)$$

2.4.2 Anisotropy Measures

Invariant measures of the diffusivity anisotropy are made of combinations of the tensor eigenvalues. The most common measures are relative anisotropy (RA), fractional anisotropy (FA), and volume ratio (VR), defined as:

$$\begin{aligned} RA &= \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{2MD}} \\ &= \frac{\sqrt{3} \|\mathbf{D} - Tr(\mathbf{D})\mathbf{I}/3\|}{\sqrt{2} Tr(\mathbf{D})} \end{aligned} \quad (2.16)$$

$$\begin{aligned} FA &= \frac{1}{\sqrt{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \\ &= \frac{\sqrt{3} \|\mathbf{D} - Tr(\mathbf{D})\mathbf{I}/3\|}{\sqrt{2} |\mathbf{D}|} \end{aligned} \quad (2.17)$$

and

$$\begin{aligned} VR &= \frac{\lambda_1 \lambda_2 \lambda_3}{MD^3} \\ &= \frac{|\mathbf{D}|}{MD^3}, \end{aligned} \quad (2.18)$$

where $|\cdot|$ denotes the Frobenius or tensor norm.

One advantage of these measures is that they can be calculated without first calculating the eigenvalues of the tensor. As discussed in Section 2.1, myelination diseases can cause a drop in the diffusion anisotropy. Hence, the above anisotropy measures, especially FA (see Figure 2-4), are sometimes used in clinical studies to diagnose diseases and assess the progress of the therapy.

2.4.3 Orientation Mapping

The idea that the main orientation of the diffusivity can be represented by the eigenvector that corresponds to the largest eigenvalue provides a powerful tool to visualize the organization of the fiber tracts. Various methods have been proposed to display

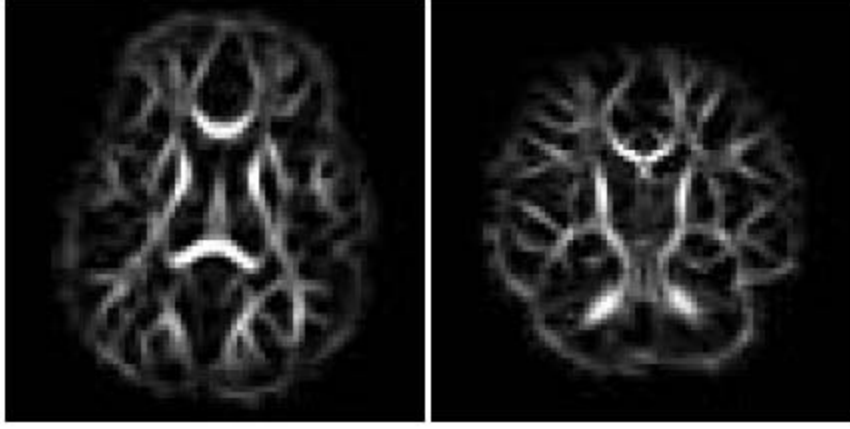


Figure 2-4: The fractional anisotropy map in axial and coronal views. The bright regions (higher FA value) represent the regions with dense fibrous tissue as the water diffusivity is highly restricted in the one direction that is parallel to the orientation of fiber tracts in the brain.

the orientation on a voxel-by-voxel basis; color maps, ellipsoids, octahedra, and quiver representations are among the most common approaches.

2.4.4 Geometrical Anisotropy Measures

Assuming that the three eigenvalues of the tensor are $\lambda_1 \geq \lambda_2 \geq \lambda_3$, Westin *et al.* showed that the tensor can be decomposed into three components of linear, planar, and spherical diffusivity, \mathbf{D}_l , \mathbf{D}_p , and \mathbf{D}_s , respectively [97]:

$$\mathbf{D} = (\lambda_1 - \lambda_2)\mathbf{D}_l + (\lambda_2 - \lambda_3)\mathbf{D}_p + \lambda_3\mathbf{D}_s \quad (2.19)$$

The following shape measures can then be defined by normalizing to the tensor's norm:

$$c_l = \frac{\lambda_1 - \lambda_2}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (2.20)$$

$$c_p = \frac{2(\lambda_2 - \lambda_3)}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (2.21)$$

$$c_s = \frac{3\lambda_3}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (2.22)$$

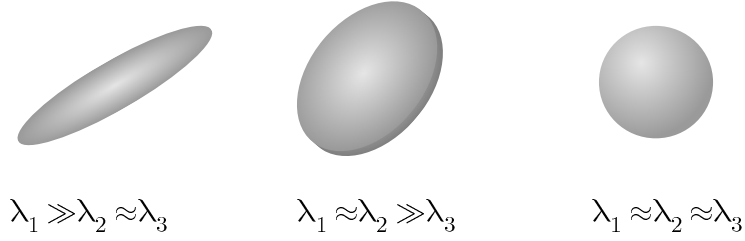


Figure 2-5: Ellipsoidal representation of a tensor with either linear, planar, or spherical symmetry.

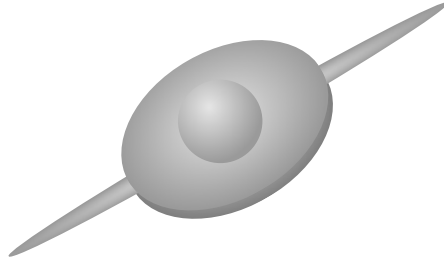


Figure 2-6: A composite representation of a tensor decomposed to its linear, planar, and spherical components.

which are linear, planar, and spherical measures, respectively. Figure 2-5 shows how the ellipsoid representation of the tensor depends on the relative magnitude of the eigenvalues. In the first case, λ_1 is significantly larger than the other two eigenvalues. The tensor is represented with a prolate ellipsoid, and $c_l \gg c_p, c_s$. In the second case, λ_1 and λ_2 have comparable magnitude, but λ_3 is significantly smaller. This is represented by a planar ellipsoid characterized by a large c_p . Finally, when all eigenvalues have comparable magnitude, the tensor can be represented by a sphere and c_s prevails. The above decomposition of the tensor provides a new visualization method as depicted in Figure 2-6.

2.5 Tractography

Tractography is a popular procedure for constructing and visualizing the underlying fiber organization by following the local orientation of the diffusivity tensor at each point (See Fig. 2-7). Throughout this thesis, we call each of the paths generated



Figure 2-7: Example of fiber trajectories extracted using a tractography algorithm on the whole human brain. Tractography seed locations, colors, and transparencies were manually chosen by a neuroanatomist. [63]

by a tractography algorithm a trajectory. Assuming that the fiber trajectory can be described by a 3D curve, $\mathbf{r}(s)$, parametrized by the arc length, s , the evolution of the curve is described by the Frenet's equation:

$$\frac{d\mathbf{r}(s)}{ds} = \mathbf{t}(s) \quad (2.23)$$

where $\mathbf{t}(s)$ is the unit tangent vector. The key idea of the tractography algorithms is to determine the tangent vector as a function of the local diffusivity at each point so that it follows the orientation of the tensor and then solve the above differential equation numerically with an initial condition $\mathbf{r}(0) = \mathbf{r}_0$, which specifies the starting (seed) point of the trajectory. The simplest way to integrate the above differential equation is by using Euler's method [8]:

$$\mathbf{r}(s + \Delta s) = \mathbf{r}(s) + \Delta \mathbf{t}(s) \quad (2.24)$$

While Euler's method is very easy to implement, it is susceptible to error accumu-

lation and instability. Second-order or adaptive fourth-order Runge-Kutta methods have been used by several groups to overcome these problems [8].

A significant amount of work has been invested in tractography in recent years. As stated earlier, the focus of this thesis is not on tractography itself, but on how to process and to extract information from the resulting trajectories, therefore we are content to describe three major approaches and their properties.

2.5.1 Streamline Tractography

In regions with high diffusion anisotropy it is plausible to assume that the eigenvector that corresponds to the largest eigenvalue of the diffusivity tensor is parallel to the local orientation of the fiber tracts. Streamline tractography methods are based on this assumption, i.e.,

$$\mathbf{t}(s) = \hat{\mathbf{e}}_1(\mathbf{r}(s)), \quad (2.25)$$

where $\hat{\mathbf{e}}_1(\mathbf{r}(s))$ is the unit eigenvector that corresponds to the largest eigenvalue of the tensor at $\mathbf{r}(s)$ [8].

The main disadvantage of streamline tractography is that if the evolving trajectory reaches a point in the space where anisotropy is small, further evolution of the trajectory is susceptible to numerical errors in determining the largest eigenvalue and the corresponding eigenvector. Usually a threshold on the minimum FA is specified as a stopping criteria. However it is sometimes desired to bypass individual points where anisotropy is locally small due to either image imperfections and noise, or where two or more fiber tracts cross each other, for example the chiasm region in the optical nerve.

Streamline tractography has been used for all of the experiments in this thesis. The steps are summarized in Algorithm 2.1.

2.5.2 Diffusion Deflection Tractography

Diffusion deflection tractography [97] has been proposed to overcome the disadvantage of streamline tractography when regions with low anisotropy are reached. In regions

Algorithm 2.1:

Input: The diffusion tensor field and seed points.
Output: Trajectories of the fiber bundles that start from the seed points.

```

1 Take the seed points  $\{\mathbf{r}_i(0)\}_{i=1}^N$ 
2 foreach seed point do
3   repeat /* stopping criteria. */
4   |   foreach  $sign = \pm 1$  do /* move in both directions. */
5   |   |    $\mathbf{r}_i(s + \Delta s) = \mathbf{r}_i(s) + sign \Delta s \hat{\mathbf{e}}_1;$ 
6   |   |    $s = s + \Delta s;$ 
7   |   end
8   until  $FA(\mathbf{r}_i(s)) < 0.15;$ 
9 end
```

with low anisotropy, such as fiber crossing regions, the direction of fastest diffusion is not well defined. Diffusion deflection tractography uses the entire tensor to deflect the propagation direction of the trajectory, i.e.,

$$\mathbf{t}(s) = \mathbf{D}(s)\mathbf{t}(s - \Delta s) \quad (2.26)$$

The propagation direction depends on the shape of the local tensor and the direction of $\mathbf{t}(s - \Delta s)$ with respect to the tensor frame. For prolate, anisotropic tensors, with a small angle between $\mathbf{t}(s - \Delta s)$ and the main eigenvector of $\mathbf{D}(s)$ the trajectory is deflected toward this eigenvector. When anisotropy is low or when the angle between $\mathbf{t}(s - \Delta s)$ and the main eigenvector is large, the deflection is smaller. In a limiting case where there is no anisotropy the propagation direction does not change.

It has been shown that this method is less sensitive to noise and image imperfections [51] but our experiments have shown that it does not perform well on extracting curved tracts due to its tendency to preserve the previous direction at each step.

2.5.3 Stochastic Tractography

Both streamline and diffusion deflection methods are deterministic tractography approaches, i.e., the course of the extracted trajectories from a group of seed point is determined solely by the diffusion tensor data. Various stochastic tractography meth-

ods have been proposed to alleviate the shortcomings of deterministic approaches in crossing regions by adding some randomness to the deterministic tractography [16]. These methods consider a probability model on the fiber orientation, instead of using either only the major eigenvector or the full tensor. Rather than producing one trajectory from each seed point, multiple trajectories are produced by sampling the probability model. Unlike the previous methods, these methods can deal with the crossing problem as they allow branching, in the sense that many trajectories are produced on a random basis so the total output can capture a branching structure. The disadvantage of stochastic methods is the explosion of the generated data for analysis or visualization since unlike streamline tractography, when a single highly probable trajectory is generated, they can produce a huge number of trajectories, on the order of 1,000, per seed point.

2.6 Related Work on Quantitative Analysis of White Matter Fiber Tracts

A significant amount of work has been devoted to extracting information from diffusion images to study brain changes related to development [42], aging [78], and different pathologies such as multiple sclerosis [96], schizophrenia [68], Parkinson’s [79], and Alzheimer’s [15]. The ultimate goal of these studies is to understand the brain connectivity and integrity and how they are affected by neural diseases through quantitative analysis of diffusion images. To this end, many algorithms have been proposed in recent years to make such analysis possible.

To understand whether new algorithms are necessary for analysis of diffusion images, we categorize the existing algorithms based on what is measured and where it is being measured. The former determines whether specific procedures are needed to extract the desired quantitative parameters from diffusion data and to perform group averaging and comparison, while the latter tells us how to specify the location where these parameters are extracted and how to make sure that in a population

Table 2.1: Categorizing DT-MRI analysis studies based on the quantity being measured. In each case the required tools and the approaches taken to develop these tools are listed.

Measured Quantity	Required Tools	Approach
Scalar parameters (such as FA)	None	
Diffusion tensor	Methods to average and compare tensor data	Riemannian [35, 53, 70] Log-Euclidean [6] Geodesic-loxodrome [48]
Shape descriptors	Methods to calculate shape descriptors	Curvature and torsion [11, 39, 57]

study these locations correspond to each other.

The easiest to measure and the most common quantitative parameters are scalar diffusion parameters, such as FA or tensor eigenvalues. Such choices are also motivated by clinical applications where axon integrity is correlated to diffusion parameters, as discussed in Section 2.1. However, information about fiber orientation encoded in the tensor data is often lost when using scalar diffusion measures. Changes in fiber orientation can be seen, for example, in the vicinity of brain lesions. Hence, another possibility for quantitative analysis is to compare orientation information or the full tensor data. To enable this, algorithms have been proposed to average and compare tensor data [6, 35, 48, 53, 70]. Finally, the last option is to measure shape descriptors of the fiber bundles, for example curvature and torsion of the fiber trajectories [11, 39, 57]. One particular application of these measures is to study changes in the shape of the fiber tracts during brain development.

The above mentioned options and the required algorithms to deal with each option are summarized in Table 2.1.

Table 2.2 lists major options to define the location where quantitative parameters are measured, the required tools for each option, and the approaches proposed to address these requirements. Voxel-based analysis, in principle, offers an excellent spatial resolution when performing group averaging and identifying group differences, but is sensitive to alignment errors. Registration algorithms are needed to map all

Table 2.2: Categorization of DT-MRI analysis work based on where the quantitative parameters are measured and compared. In each case the required tools specifying the location and the approaches proposed to address these requirements are listed.

Measurement location	Required Tools	Approach
Voxel-based	DT-MRI registration	Diffusion scalars [75] Multi-channel (Orientation) [74] Tensor [2, 20, 21, 72, 100] Tract-based [82, 99, 102]
ROI-based	ROI specification	Scalar segmentation Tensor segmentation [45, 54, 76, 93] Tractography [46, 47, 69]
Tract-oriented	Tractography	[10, 38, 97]
	Grouping (Clustering)	Unsupervised [17, 64, 81, 88] Knowledge-based [41, 58], [60, 61] (This thesis)
	Point-Correspondence	Land marks and Procrustes alg. [11, 22] Curve matching [30, 56, 66] Distance map [59](This thesis)

subjects into a common space. Methods proposed to register conventional MRI can be easily adopted by using scalar diffusion fields, such as FA or ADC. However, it has been shown that the use of the directional information present in the diffusion tensor elements significantly improves the accuracy of the registration results [74]. Algorithms have also been proposed to use the whole tensor field for registration [2, 21, 72, 100], which in turn requires tensor re-orientation [2, 67]. Apart from the sensitivity of the voxel-based method to registration process, the spatial resolution of the method is often lost due to practical considerations [47]. The data is often over-smoothed to ensure robust registration and significance of the statistical analysis. Hence, it is difficult to establish correspondence between the voxel regions where significant group difference has been observed and anatomical structures. Alternative approaches include [82], where the skeleton of the FA volumes is extracted and aligned to compute the registration matrix. Similar approaches have been proposed in [102], where fiber trajectories are aligned to an atlas to obtain the transform parameters and in [99], where the medial surface of fiber tracts is used for registration.

ROI-based methods are currently the most popular approach in quantitative analysis of MRI. To avoid the manual delineation of ROIs and gain reliability and time efficiency, algorithms have been proposed to segment the diffusion images, mostly from the FA volumes and recently using the whole tensor information [45, 54, 76, 93]. Tract-based approaches [46, 47, 69] have also been proposed in which the trajectories extracted by a tractography algorithm form the ROI. Regardless of how the ROIs are specified, local variation of the quantitative parameters is not captured in such analysis.

In this thesis we focus on tract-oriented quantitative analysis, in the sense that we are interested in the variation of diffusion parameters or shape descriptors along the extracted trajectories that belong to an anatomical tract. Analogous to segmentation in ROI-based analysis, clustering approaches have been proposed to group the trajectories into bundles. Clustering ensures that the quantitative parameters are measured only on trajectories that belong to a bundle. Hence, one would compare the same tracts, when performing a group analysis and the observed differences will be due to differences in the properties of the specific tracts rather than differences in the overall anatomy/shape of the individual brains. In general, these algorithms all share the common theme of first defining a similarity metric between the trajectories, and then employing an algorithm for clustering based on the established similarity measure [88].

The output of clustering methods is a set of labeled trajectories, each assigned (probabilistically) to a cluster. In most studies, point-by-point correspondence between the trajectories of each cluster, however, is not determined rigorously. In one of the early works on DT-MRI analysis, Ding, *et al.* [30] tackled the issue of quantification of tracts by finding the corresponding segments, which they defined as the portion of a trajectory that has point-wise correspondence to a portion of another trajectory. They assumed that the seed points of the two trajectories to be compared correspond to each other, which is not the case unless all trajectories are seeded from a small ROI. The algorithm is thus inadequate for whole brain fiber analysis.

Batchelor, *et al.* [11] proposed different tools to quantify the shape of fiber tracts.

They noted the problem of point correspondence but made the assumption that it is approximately achieved by proper choice of the seed point and regularly sampling the arc-length. With point correspondence roughly known, they applied the Procrustes algorithm to register the trajectories. There is also a series of work by Gerig *et al.* (c.f. [39]), which described methods and applications for tract-oriented quantitative analysis. They dealt with the issue of correspondence by letting the user define a common origin for the set of trajectories in each cluster, based on geometric criteria or based on anatomical landmarks. In their latest work [22], they also proposed the Procrustes algorithm for the registration of the trajectories to compute the average tensor. Although their approach provides some valuable information about the fibers, its applicability is limited by the need for manual intervention to set common start points for all the trajectories in a cluster. Also, they assume all the trajectories in a cluster have the same length, which is a reasonable assumption only if a small ROI is considered for the tractography seed points and they end roughly in a common area. Otherwise a thorough preprocessing is required.

In one of our earlier works [56], we used a string matching algorithm to align all extracted trajectories with each cluster center at each iteration of our expectation-maximization clustering. The accuracy of this approach was limited by the simple curve matching algorithm used. In subsequent work [59], we presented the idea of constructing a Voronoi diagram by building a distance map (as detailed in Chapter 3 of this thesis) to find the closest points on the cluster centers to the points on the trajectories and so establishing the point correspondence in an efficient way. In another work [66], a common coordinate system was constructed in a similar fashion by matching each point on the prototype fiber (cluster center) to the closest point(s) on other trajectories.

Most clustering approaches used to group fiber trajectories are unsupervised [17, 64, 81], and hence do not necessarily produce anatomically-meaningful clusters. We believe that a supervised clustering, which benefits from anatomical information, not only produces anatomically meaningful clusters, but also yields more robust results that are less sensitive to the presence of outliers and imperfections in the DT-MRI

data. It also facilitates the population studies by making the correspondence between the bundles across subjects.

Based on the discussion above, the elements of tract-oriented analysis (as defined here) are: (a) tractography, where many methods have been already proposed; (b) clustering, where to our knowledge no method has been proposed in the past to incorporate anatomical information rigorously; and (c) determining point correspondence, where proposed methods are in general too simplistic. The rest of this thesis provides efficient methods to determine the point correspondence between trajectories and to perform the clustering in a Bayesian framework, where anatomical information can be used rigorously.

Chapter 3

Point Correspondence and Similarity Measures for 3-D Curves

As discussed in the previous chapter, clustering fiber trajectories into anatomically known bundles and determining their point correspondence are two integral parts of the tract-oriented quantitative analysis. The former requires a similarity measure of 3-D curves to discriminate the trajectories. Determining the point correspondence between a pair of trajectories is not only essential for tract-oriented analysis, but also reduces the calculation of the similarity measure to the calculation of the Euclidean distance between corresponding points.

Determining the point correspondence between 3-D curves is not a trivial task. Although many authors acknowledge that point-by-point correspondence of the trajectories should be defined by a curve matching algorithm for accurate clustering and quantitative analysis [11, 30, 39], to our knowledge this problem has not been solved to date. The difficulty lies in the fact that the number of trajectories is usually very large, especially when the analysis is performed on a population or on the whole brain. This makes it computationally inefficient, if not impossible, to perform a rigorous curve matching algorithm on every pair of trajectories. In this chapter, we first review popular methods for calculating the distance between curves, and then we list the measures that have been used in DT-MRI analysis to date. We next propose a novel approach that determines the point correspondence and defines the similarity

measure at the same time and is computationally efficient and sufficiently robust to be used for large groups of trajectories.

3.1 Similarity Measures for Curves

Many similarity measures are based on the L_p distance or the Minkowski distance between two points [91]:

$$L_p(\mathbf{x}, \mathbf{y}) = \left(\sum_j |x_j - y_j|^p \right)^{1/p} \quad (3.1)$$

For $p = 2$ this yields the Euclidean distance and for $p = 1$ it reduces to the Manhattan distance.

Different distance measures have been defined to show the dissimilarity between two groups of points. This section reviews the most common choices.

3.1.1 Bottleneck Distance

For two point sets of equal size, the bottleneck distance is defined as the minimum of the maximum point-to-point distance taken over all one-to-one correspondences between the two sets [43]. Variations on the bottleneck distance are the minimum weight distance, the most uniform distance, and the minimum deviation distance. However, in many applications, including ours, the two point sets (3-D curves in our work) do not have the same size.

3.1.2 Hausdorff Distance

The directed Hausdorff distance from one set of points to another set is defined as the lowest upper bound (supremum) of the distance to the second set over all points in the first set. The Hausdorff distance between the two sets is defined as the maximum of the two directed Hausdorff distances.

The Hausdorff distance is very sensitive to noise: a single outlier can determine the distance value [91]. For finite point sets, a similar measure that is not as sensitive

is the partial Hausdorff distance. It is the maximum of the two directed partial Hausdorff distances, defined as the k -th value in increasing order of the distance from a point in the first set to the second set. Alternatively, the p -th order mean Hausdorff distance has been defined, which is less sensitive to noise, but still not very robust [91].

3.1.3 Fréchet Distance

The Hausdorff distance does not consider the order of the points on the curve; for all points on one curve the closest distance to the other curve is measured. However, if we move along both curves simultaneously and measure the distance between the corresponding points, the maximum of these distances may be larger. The Fréchet distance takes care of the order of the points on the curves, by finding the maximum distance between the two curves for each monotonic point correspondence and taking the minimum over all point correspondences. Algorithms have been proposed to speed up the search process, but this metric is still not easy to implement [3].

3.1.4 Nonlinear Elastic Matching Distance

For each pair of corresponding points taken from the two curves, the stretch is set equal to 1 if either of the points have correspondence to the preceding point on the other curves as well, and otherwise equal to zero. The nonlinear elastic matching distance is the minimum over all correspondences of the sum of the stretches plus the difference between the tangent angles at each pair of corresponding points [24]. It should be noted that the nonlinear elastic matching distance is not a metric. The metric proposed in our work uses a concept similar to the stretch function we propose in this chapter.

3.1.5 Turning Function Distance

For a 2-D curve, the cumulative angle function, or turning function gives the angle between the counterclockwise tangent and the x -axis as a function of the arc length. This function is invariant under translation, while rotation results in a shift in the

function. The distance between two 2-D curves can be defined as the integral of the L_p distance between their turning function representation [5]. Extension of turning function distance to 3-D curves is not straightforward. Moreover, in our application the spatial distance between two curves is also important, so a translation-invariant representation is not a good choice.

3.1.6 Reflection Distance

The reflection metric is an affine-invariant metric defined on finite unions of curves on a plane or surface [91]. For each point on the plane the *visibility star* is defined as the union of open line segments connecting points of the curve that are visible from that point. The *reflection star* is defined by intersecting the visibility star and its reflection with respect to the point. The reflection distance is calculated by integrating the difference between the area of the reflection stars defined at each point on the plane and with respect to the two curves. This metric is robust to deformation, blur, crack, and noise occlusion, but computing the distance by explicitly constructing the visibility stars is highly time consuming.

3.1.7 Transport Distance

The transport distance between two curves is the minimum amount of work needed to transform one curve to the other [91]. The flow between each pair of points from the curves is defined to satisfy several properties: (a) All flows are non-negative; (b) For each point on the second curve, the inflow, i.e., sum of the flow from all points on the first curve to that particular point, is not greater than one; (c) Similarly, the out-flow from each point on the first curve is not greater than one; (d) The sum of all flows is equal to the minimum number of points on the curves. The transport distance is calculated by multiplying the flow by the distance between each pair of points, summing up over all pairs, and finding the minimum possible over all choices of the flow matrix.

3.2 Similarity Measures Used in DTI Analysis

To get around the difficulty of performing a sophisticated curve matching algorithm, many authors attempt to define simple, yet reasonable similarity measures. For example, Brun *et al.* simply used the start and end points of the trajectories for clustering [16] as their model is to match points on the cortex, and hence if connected, it is less important to know exactly which path the tract took. In [17], they used a 9-D shape descriptor for each trajectory, including the mean and square root of the covariance matrix of its points and then computed the pairwise Euclidean distance of the descriptors in a spectral clustering scheme. The clustering proposed in [30] defines the similarity measure based on the length ratio and the Euclidean distance between the corresponding segments of each pair of trajectories. However, the corresponding segments are determined with the assumption that the seed points of all the trajectories correspond to each other. In [81] a fuzzy c-means clustering algorithm is proposed and various distance measures between the trajectories are used, including the dot product of the corresponding tangents of the trajectories and the average distance between points along them. Corouge, *et al.*, [23] implemented several distance measures such as closest point distance, mean distance of closest points (symmetric chamfer distance), and the symmetric Hausdorff distance, and showed that each distance metric has certain advantages and shortcomings. In a spectral clustering approach, O'Donnel *et al.* [64] build an affinity matrix upon pairwise symmetric Hausdorff distances between trajectories. Leemans *et al.* [52] find the closest subcurves in the curvature-torsion space and vary the length of the subcurves to deal with the curve matching problem. They further use a scale-space method to construct the space curves at different levels of detail. The spatial information is, however, not utilized in their approach. Furthermore, the time-efficiency of their method for whole brain clustering or population study is questionable as the complexity of their approach grows with the square of the number of subcurves. In [58], we used a B-spline representations of trajectories to compare the trajectories extracted from the subject to those from an atlas.

From the discussion above, it is clear that there is no consensus in the DTI community about the best similarity measure to compare fiber trajectories. The Hausdorff distance perhaps is the most popular measure that is adequate to capture the local relationship between two curves, however it is very sensitive to noise in a sense that a single outlier can determine the distance value [91].

As stated earlier, apart from defining a similarity measure needed for clustering, we also need to establish point correspondence between a pair of trajectories to perform quantitative analysis along the fibers. We reviewed the related work on the calculation of point-correspondence in the Table 2.2 in Chapter 2. In this chapter, we present our novel approach for a stable and computationally efficient way to define similarity and point-correspondence between the trajectories using the distance transform.

3.3 Our Proposed Approach: Distance Transforms

We treat each trajectory as a sampled 3-D curve, i.e., an ordered set of points, so the i th trajectory is represented as $\mathbf{r}_i = \{\mathbf{r}_{ij}\}$. As will be outlined in the next chapter, the set of trajectories is clustered into a number of subsets by assigning a membership probability p_{ik} to each trajectory, \mathbf{r}_i , to denote its membership in the k th cluster ($\forall i, \sum_k p_{ik} = 1$). For each cluster, a 3-D curve, $\boldsymbol{\mu}_k = \{\boldsymbol{\mu}_{kj}\}$, is defined as the cluster center (prototype trajectory) where each point, $\boldsymbol{\mu}_{kj}$, is obtained as the average of all of its corresponding points on the trajectories: $\boldsymbol{\mu}_k = \sum_i p_{ik} \mathbf{r}_i^{(k)}$, where $\mathbf{r}_i^{(k)}$ is the trajectory \mathbf{r}_i , re-parametrized to have point correspondence to cluster k , and the summation is performed over all trajectories (See Figure 3-1).

Our space includes a set of 3-D curves and a number of cluster centers. From each center, $\boldsymbol{\mu}_k$, we construct a Euclidean distance map, that is, for each point in the space, \mathbf{x} , its distance is given by:

$$\mathcal{D}_k(\mathbf{x}) = \min_j d(\mathbf{x}, \boldsymbol{\mu}_{kj}) \quad (3.2)$$

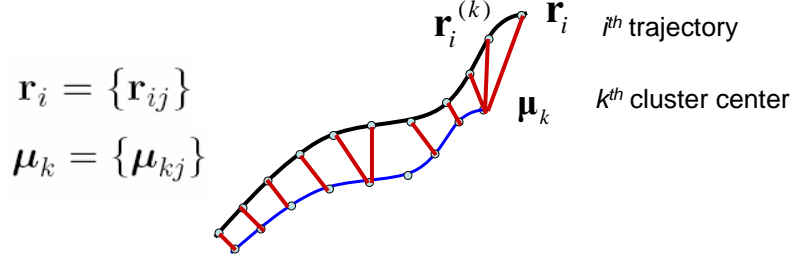


Figure 3-1: A schematic representation of a trajectory (\mathbf{r}_{ik}) and a cluster center ($\boldsymbol{\mu}_k$). They are equidistance sampled. The \mathbf{r}_{ik}^k represents the trajectory with its point-correspondence to cluster k is determined.

and the nearest-neighbor transform, \mathcal{L}_k , for each point in the space, is given by:

$$\mathcal{L}_k(\mathbf{x}) = \arg \min_j d(\mathbf{x}, \boldsymbol{\mu}_{kj}) \quad (3.3)$$

where $d(\mathbf{x}, \boldsymbol{\mu}_{kj})$ is the Euclidean distance from the point \mathbf{x} in the space to the j th point on the k th center. Each element of \mathcal{L}_k will thus contain the linear index of the nearest point of the center $\boldsymbol{\mu}_k$. Figure 3-2 shows the distance map and label map constructed from a sample 2-D curve. The label map partitions the space into Voronoi cells that each correspond to a point of the center. Now, for every curve, $\mathbf{r}_i = \{\mathbf{r}_{ij}\}$ in the space, the distance to the center $\boldsymbol{\mu}_k$ can be measured simply as:

$$d_E(\mathbf{r}_i, \boldsymbol{\mu}_k) = \sum_j \mathcal{D}_k(\mathbf{r}_{ij}), \quad (3.4)$$

and by projecting the curve onto the label map, its point correspondence to the center is readily achieved (See Figure 3-2).

We also note that a spatial distance by itself does not encode enough information for measuring the pair-wise similarity. One obvious issue is the variable lengths of the trajectories. Two examples are shown in the Figure 3-3: (a) shows the case where the trajectory (marked by blue dots) has a very small distance to the cluster center (marked by red dots), but is not similar in shape to it. So we want some penalty to account for this shape dissimilarity. (b) shows another case in which the trajectory has roughly the same distance to either of the cluster centers, but we prefer it to be

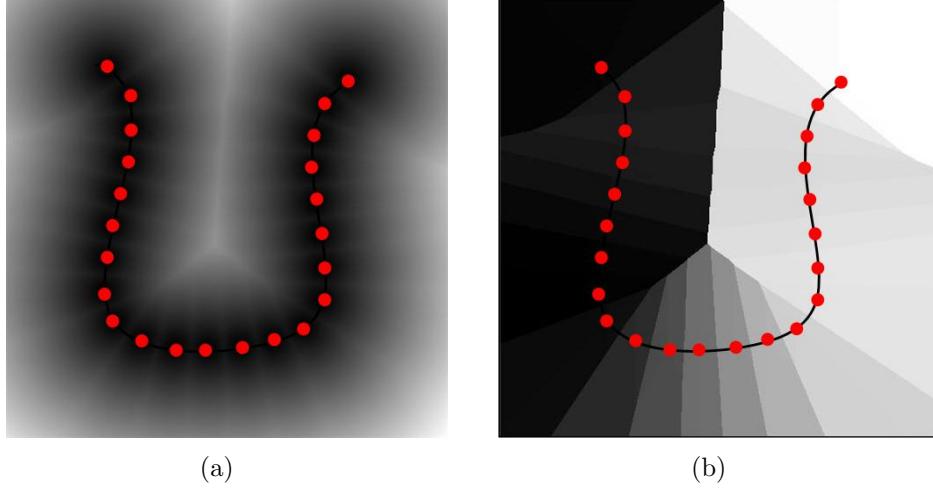


Figure 3-2: (a) Distance map from sample points on a cluster center and (b) the point correspondence label map with the center overlaid. Each region in the label map, displayed by a different color, consists of all of the points in the space that have the minimum distance to a specific point on the cluster center. Therefore, projecting any curve onto this label map determines the point correspondence of each of its samples to the center based on which region that sample is located.

assigned to the left one.

Given that we have the knowledge of point-correspondence, we add a penalty for each missing point, when a trajectory is shorter than the cluster center, as shown in Figure 3-4(a). Another issue is whether the trajectory has one-to-one point correspondence to the cluster center, which is the case when they are similar in shape. Thus, any repeated or missing match represents shape dissimilarity. A penalty is needed to be added if there are multiple points on the trajectory that correspond to a single point on the cluster center, as in Figure 3-4(b). However, if multiple correspondence is merely because the trajectory is longer than the cluster center no penalty is added, since the trajectory is already penalized, as shown in Figure 3-4(c). We add the penalty term, d_{penalty} as described above to the Euclidean distance and normalize to the length of the fiber trajectory, L_i . The normalization is to make the distance independent of length of the trajectory. We denote the adjusted distance by $d_a(\mathbf{r}_i, \boldsymbol{\mu}_k)$:

$$d_{ik} = d_a(\mathbf{r}_i, \boldsymbol{\mu}_k) = \frac{d_E(\mathbf{r}_i, \boldsymbol{\mu}_k) + d_{\text{penalty}}(\mathbf{r}_i, \boldsymbol{\mu}_k)}{L_i}. \quad (3.5)$$

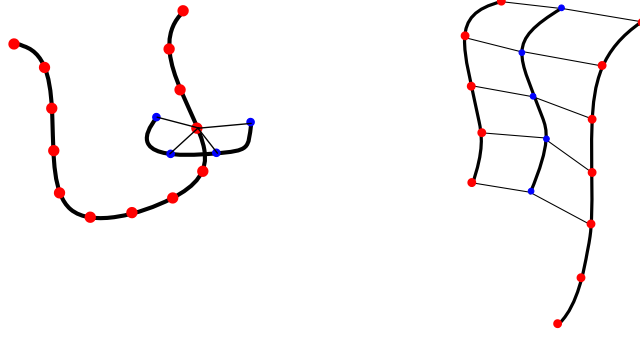


Figure 3-3: Schematic representation of cases where a penalty for shape dissimilarity is required. (a) A trajectory (marked by blue dots) that has a very small distance to the cluster center (marked by red dots), but is not similar in shape to it. (b) A trajectory that has roughly the same distance to either of the cluster centers, but we prefer it to be assigned to the left cluster center that has roughly the same length as the trajectory.

where d_{penalty} is equal to the number of missing points times the average distance of the matched points of the trajectory from the cluster center, d_{av} . An efficient way of implementing this in the algorithm is to set

$$d_{\text{penalty}}(\mathbf{r}_i, \boldsymbol{\mu}_k) = ((L_k - L_i) + \text{Number of repeated matches}) \times d_{av} \quad (3.6)$$

where L_k and L_i are the length of the cluster center and the trajectory, respectively.

Note that this penalty term is similar to the stretch term used in the nonlinear elastic matching distance, described earlier.

As an optional modification of the method, one can model the spread of the trajectories in each cluster, by associating a covariance matrix with each point on the cluster center ¹. We denote the 3×3 covariance matrix attributed to the point $\boldsymbol{\mu}_{kj}$ on the k th cluster with $\boldsymbol{\Lambda}_{kj}$. When computing the similarity of each trajectory to the center, the variability of the covariance matrix along the center should be taken into account, so that the trajectory points associated with a portion of the cluster center that has larger covariance are penalized less. This is done by computing the

¹The computation of covariance matrix is discussed in Chapter 4

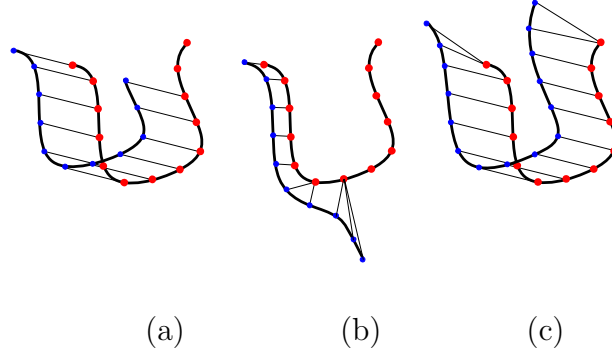


Figure 3-4: Adjustments made to the total distance between the trajectory (blue) and the cluster center (red) calculated by Eq. (3.7). The penalty is equal to the number of missing point times the average distance of the matched points, d_{av} . This is implemented more efficiently as $d_{\text{penalty}} = ((L_k - L_i) + \text{Number of repeated points}) \times d_{av}$, where L_k and L_i are the length of the cluster center and the trajectory, respectively. (a) $d_{\text{penalty}} = 2d_{av}$, (b) $d_{\text{penalty}} = 6d_{av}$, (c) $d_{\text{penalty}} = 0$.

Mahalanobis distance:

$$d_M(\mathbf{r}_i, \boldsymbol{\mu}_k) = \sqrt{\sum_{j=1}^{L_i} (\mathbf{r}_{ij}^{(k)} - \boldsymbol{\mu}_{kj})^T \boldsymbol{\Lambda}_{kj}^{-1} (\mathbf{r}_{ij}^{(k)} - \boldsymbol{\mu}_{kj})} \quad (3.7)$$

3.4 Discussion

As stated earlier, there is no fundamentally correct way to quantify the similarity between two curves and hence it is not possible to mathematically assess the quality of the distance measures introduced in this chapter. In Table 3.1, we qualitatively compare our introduced distance measure with our judgment on what is expected from a distance measure in our application. Comparison is also made with both unidirectional and bidirectional Hausdorff distances. Note that in this comparison, both the cluster center (marked with red dots) and the trajectory (marked with blue dots) are uniformly sampled - with the same down-sampling ratio.

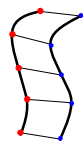
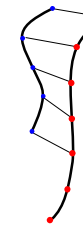
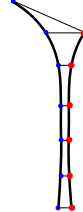
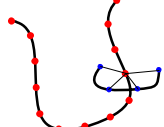
The first row in the table shows a pairing of trajectory-cluster center which is very similar in shape. Assuming that the average distance between the corresponding points is equal to unity, one expects the distance between the two to be equal to one. In fact, all three distance measures yield a distance of one. The second row

shows a pair where the trajectory is smaller than the center. Hence, once the point correspondence from the trajectory to the center is determined, there are two points on the center that do not have any corresponding points on the trajectory (two missing points). Based on our judgment, the distance between the trajectory and the center must be somewhat more than unity, but less than two, i.e., less than the case where there is no missing point but the average distance of the corresponding points is two. The penalty term in our distance measure satisfies this requirement. The third row shows a case of the trajectory that is very close to the center for most of its length, but fans out at one end. Such cases might occur, for example, in corticospinal tracts. Both Hausdorff distances give a value of 2, while our measure gives a distance less than one. Finally, the last row shows a case where the trajectory and center are not similar at all, in spite of the fact that overall the trajectory is located close to the center. The unidirectional Hausdorff distance clearly fails in this case, while both bidirectional Hausdorff and our measure result in relatively high numbers, in agreement with the fact that the two curves are not similar at all.

3.5 Summary

In this chapter we addressed the problem of defining a similarity measure between 3-D curves as a prerequisite of the clustering step. Since in our application we mainly deal with a large number of trajectories (3-D curves), a pairwise computation of a similarity measure and invoking a curve matching algorithm for finding the point correspondences is not a computationally efficient method. In this chapter we took advantage of the fact that although the number of trajectories is large, the number of desired clusters is usually small. We therefore reduced the problem to finding the distance and point correspondence of each trajectory to each cluster center (a prototype trajectory). We proposed a novel method for measuring similarity and establishing point correspondences between the trajectories and each of the cluster centers by building a distance map and the corresponding Voronoi diagram for each center. We demonstrated the computational efficiency of our approach in theory and

Table 3.1: Comparison of the unidirectional and bidirectional Hausdorff distance with the distance measure introduced in this work for typical trajectory-center pairs. Blue points represent samples of the trajectory, while red points are samples of the cluster center.

Trajectory-Center Pair	Unidirectional Hausdorff	Bidirectional Hausdorff	Our Measure	Expected Dissimilarity
	1	1	1	1
	1	2	$\frac{5+2}{5} = 1.4$	$1 < d < 2$
	2	2	$\frac{4+0.6}{7} = 0.65$	$d \approx 1$
	1	4	$\frac{4+12}{4} = 4$	$d \gg 1$

practice and qualitatively compared it with the most popular methods. In the next chapter we present our mixture model clustering approach, in which the similarity measure defined in this chapter is used and the cluster centers are iteratively updated.

Chapter 4

Probabilistic Clustering of White Matter Trajectories

In this chapter we propose a novel model-based method to cluster fiber trajectories into groups on the basis of shape similarity and spatial proximity. We first briefly describe possible clustering approaches and justify our choice of mixture model clustering. The merit of such a model-based approach is contingent upon choosing an accurate probability model, for which we propose a gamma distribution over the distances between the trajectories and cluster centers. Since in the next chapter we add additional layers of probability models to incorporate anatomical knowledge, we will begin with the use of graphical models to better understand the inference problem and the relation between the variables involved. We then move to this chapter's inference problem, in which the gamma mixture parameters and therefore cluster assignments are inferred from the observed data, i.e., the distance between the trajectories and the cluster centers. The Expectation-Maximization (EM) algorithm is a powerful approach for inferring unknown parameters. We provide an introduction to the EM algorithm, followed by our clustering problem and the EM formulation used to solve it. We conclude this chapter by presenting the results of applying the EM method on some experimental data.

4.1 Desired Features and Possible Solutions

As stated in Chapter 1, the two main features, which we desired in our application, are probabilistic assignment of the trajectories to clusters and the flexibility to include anatomical information in the clustering process. To understand which clustering approaches can be adopted to satisfy these requirements, we first review the existing methods.

Clustering methods can be very broadly categorized in different ways:

- Discriminative versus generative
- Deterministic versus probabilistic
- Partitional versus hierarchical

In discriminative or distance/similarity-based methods, such as the spectral clustering approach, a distance or similarity measure between each pair of data points is first calculated and then the data is clustered based on the similarity measure. For data points that can be represented by a vector, i.e., when they have the same length and correspondence between their components is predetermined, the most common distance measures are Euclidean distance and Mahalanobis distance. For more complex data types, for example variable-length sequences, a good similarity measure is often application-dependent as discussed in the previous chapter. For large data sets, such as the trajectories in our problem, pair-wise calculation of the similarity can be computationally inefficient.

Generative or model-based methods, on the other hand, learn generative models from the data, where each model represents a cluster. Contingent on proper selection of the model, model parameters can correspond to meaningful parameters. For example, in Gaussian mixture model clustering, model parameters represent the average and spread of the data points in each cluster, while mixing weights give the prevalence of each cluster. Naturally, it is straightforward to put some limits on each of these parameters if prior information is available. In such approaches, the identification of outliers is also straightforward.

Partitional clustering methods (including mixture models) attempt to directly divide the data set into a given number of clusters. Hierarchical methods, on the other hand, proceed successively by either merging smaller clusters into larger ones (bottom-up), or by splitting larger clusters into smaller ones (top-down). These algorithms give a tree of clusters, called a dendrogram, which shows how clusters are related. The user can decide where to cut the dendrogram, and thus the algorithm is immune to over- or under- clustering, which is a common pitfall for partitional clustering methods. In our application, however, the user specifies the initial cluster centers, which are representative trajectories for each cluster, based on some anatomical information and the desired degree of detail¹. Hence, a partitional clustering method is also able to find the desired clusters.

Clustering methods can also be categorized into deterministic and probabilistic approaches. Deterministic clustering algorithms, such as k-means, assign each data point to one and only one cluster. This may lead to biased estimators of the cluster centers if the clusters overlap. A probabilistic clustering, on the contrary, deals with the inherent uncertainty in assigning the data points to clusters. More importantly, quantitative parameters can be estimated by a weighted averaging over cluster members and thus more robust results, less sensitive to the presence of outliers, may be obtained.

As stated earlier, another desired feature in our application is the ability to include anatomical knowledge in clustering. A Bayesian approach is a principled way of combining prior knowledge that comes from an atlas in our application and the information from the observed data (distances here). Our proposed model-based approach has the flexibility to benefit from prior information in different levels. Our implementation of the clustering algorithm that includes spatial prior information is discussed in the next chapter.

Based on the above discussion we opt to use mixture-model clustering in a Bayesian

¹Note that specifying the initial cluster centers by the user is not a difficult task. As will be shown later in this chapter, the final clustering results are not very sensitive to the initial cluster centers as long as these centers represent the desired clusters. So, the user can for example pick the trajectories that s/he feels are representatives of the clusters. Alternatively, these initial cluster centers can be supplied by an Atlas if such data is available.

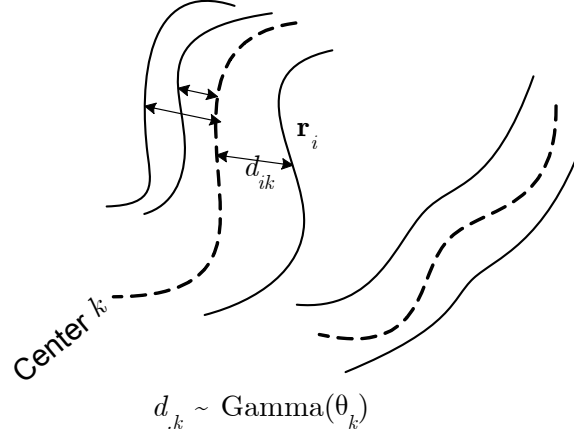


Figure 4-1: In our approach the probabilistic model is constructed on the distances between the trajectories (solid curves) and the cluster centers (dashed curves), not on the point coordinates. For each pair of trajectory \mathbf{r}_i and cluster center $\boldsymbol{\mu}_k$, a distance d_{ik} is calculated as explained in Chapter 3. As detailed in this chapter, we further assume that the d_{ik} 's follow a gamma distribution parametrized with θ_k , i.e., $d_{ik} \sim \text{Gamma}(\theta_k)$.

framework. In other words, the proposed clustering method will be a probabilistic, model-based approach, where anatomical knowledge is included with a Bayesian formulation. This novel approach is a central aspect of this thesis.

4.2 Probabilistic Model for Trajectories

All features of a probabilistic model-based approach are conditioned upon proper modeling of the data. This section discusses our modeling choices.

4.2.1 Density Estimation

As discussed in the previous chapter, the variable-length 3-D curves are mapped into a distance matrix, $\mathbf{d}_a(\mathbf{r}_i, \boldsymbol{\mu}_k) = \{d_{ik}\}_{N \times K}$ where N is the number of curves (trajectories) and K is a user-defined number of clusters. The ultimate goal of this chapter is to estimate the membership probability of each curve to each cluster based on the values of d_{ik} 's (See Figure 4-1). Here, we aim to construct a probabilistic model for each cluster.

We propose to perform the modeling on distances instead of the actual point coordinates. The reasons for this choice are:

- Calculating the probability density of the points on the curves is not straightforward as they are not statistically independent.
- We are dealing with variable-length data, as each trajectory has a different length which leads to non-identical dimensionality between the probability modes, if constructed in the space of point coordinates. Building the probability model based on the distance to the cluster centers, introduced in the previous chapter, deals with this problem effectively.

The Gaussian distribution is a common choice for the density functions of the data points. However, a Gaussian distribution does not accurately represent the nature of the distance of the 3-D trajectories from the cluster centers. In the simplest form, the number of possible trajectories with a given distance from the center grows with the distance, while the probability that they belong to that cluster decays exponentially. Among the well-known distributions, the Gamma distribution captures this combined behavior very well: the monomial term increases with the distance while the exponential term decays. Given that the d_{ik} 's are non-negative, we assume that distances for each cluster follow a Gamma distribution with shape and inverse scale parameters α_k and β_k , respectively:

$$\text{Gamma}(d|\alpha_k, \beta_k) = d^{\alpha_k-1} \frac{\beta_k^{\alpha_k} e^{-\beta_k d}}{\Gamma(\alpha_k)} \text{ for } d \in [0, \infty) \quad (4.1)$$

where $\Gamma(\cdot)$ is the gamma function. As can be seen in Figure (4-2), the shape of the Gamma distribution is controlled by parameter α_k , which is indicative of distance variation. For $\alpha_k = 1$ the distribution becomes an exponential distribution. When $\alpha_k > 1$, the distribution is bell-shaped and in the case of $\alpha_k < 1$, the distribution is highly skewed. The inverse scale parameter β_k controls the decay rate.

Figure 4-3 shows the histogram of distances of the trajectories that belong to the lower splenium (Figure 1.3) from the corresponding cluster center. A gamma distri-

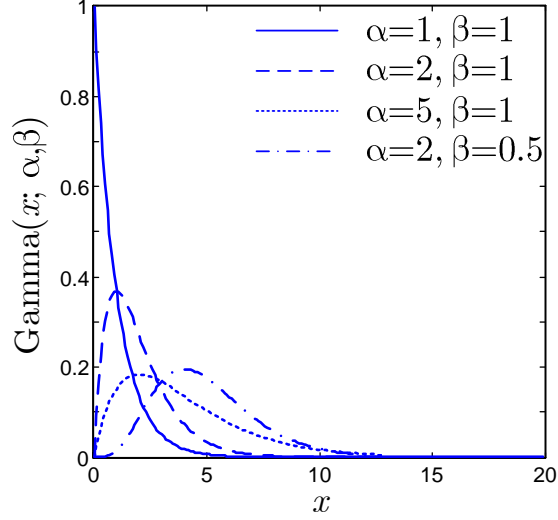


Figure 4-2: Gamma distribution function with different set of parameters that control the shape of the distribution.

bution fits well with this histogram. Note that this is just a sample, and each cluster can be fitted well with the gamma distribution with a different set of parameters.

4.2.2 Mixture Model

In mixture-model clustering, the data is modeled by a finite number of density functions, where each cluster is represented by a parametric distribution:

$$p(\mathbf{x}|\Theta) = \sum_{k=1}^K w_k f_k(\mathbf{x}|\mathbf{g}_k), \quad (4.2)$$

where \mathbf{x} is a feature vector (distances in this case), w_k 's are mixing weights, f_k is the density function of cluster k parameterized by \mathbf{g}_k , and $\Theta = \{w_k, \mathbf{g}_k\}_{k=1}^K$ is the collection of parameters that define the mixture model. We define the mixture model in the space of \mathbf{d} , as follows:

$$p(\mathbf{d}_i|\Theta) = \sum_{k=1}^K w_k f_k(\mathbf{d}_i|\mathbf{g}_k), \quad (4.3)$$

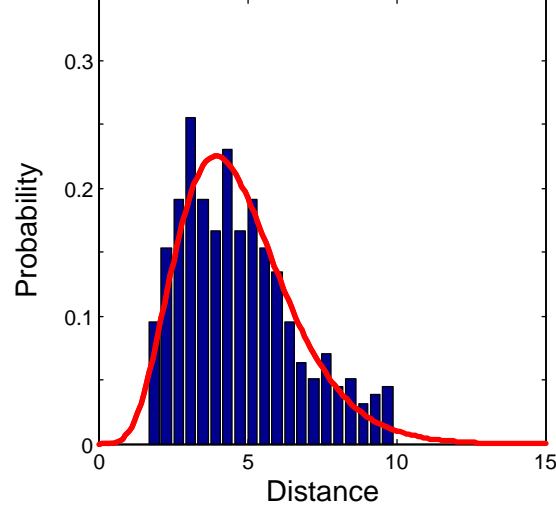


Figure 4-3: Histogram of the distances of the trajectories from the cluster center and the fitted gamma distribution.

with the density functions given by:

$$f_k(\mathbf{d}_i | \mathbf{g}_k) = \text{Gamma}(d_{ik}; \mathbf{g}_k) \prod_{j \neq k} \text{U}(d_{ij}; 0, d_0), \quad (4.4)$$

where \mathbf{g}_k is the parameter set of the Gamma distribution, $\{\alpha_k, \beta_k\}$, for cluster k , and trajectory i corresponds to cluster j , and $\text{U}(x; 0, d_0)$ is the uniform distribution function over $[0, d_0]$ with d_0 a large enough constant. The above equation means that in constructing the probability model for cluster k , we only care about the distances of trajectories to the cluster center k . The uniform distribution term is basically added to make the definition of probability model on d_i complete in K dimensions.

In the current implementation, the w_k 's are treated as unknown parameters, however they could be assumed as fixed values, taken from an anatomical prior as will be discussed in the next chapter. The goal here is to infer $\Theta = \{\mathbf{g}_k, w_k\}$ from the observed data points, \mathbf{d}_i 's.

4.3 Graphical Model for Our Clustering Problem

Graphical models help us better understand the relationship between the variables in an inference problem. Appendix A presents an introduction to graphical models.

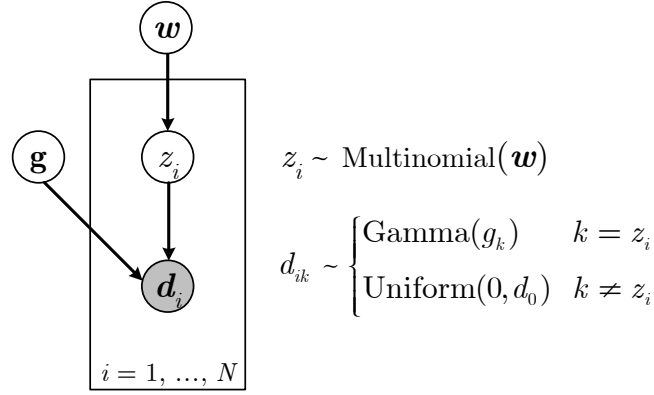


Figure 4-4: A graphical model of the clustering of fiber trajectories in the absence of an anatomical prior. The observed variables, i.e., shaded node, are the distances from each trajectory to cluster centers. z_i is defined as the label of each trajectory which takes on integer values. \mathbf{g} is the collection of the parameters for the Gamma distributions and \mathbf{w} is the mixing weights.

Figure 4-4 shows the graphical model for clustering trajectories in the absence of any anatomical spatial prior, which is the focus of this chapter. The observed variables (shaded node) are the distances of each trajectory to each cluster center. Z is an indicator variable to represent the hidden data, a collection of z_i 's, and is defined as the label of each trajectory, which takes on integer values and can be considered as samples of a single-trial multinomial distribution:

$$z_i \sim \text{Multinomial}(\mathbf{w}), \quad (4.5)$$

where $\mathbf{w} = [w_1 \dots w_K]$ are the collection of the unknown mixing parameters and K is the user-defined number of clusters. In other words, in the absence of any observation, $p(z_i = k)$ is equal to the weight of cluster k . Note that $\sum_{k=1}^K p(z_i = k) = \sum_{k=1}^K w_k = 1$.

As stated earlier, we assume that each element of the vector \mathbf{d}_i follows the distribution

$$d_{ik} \sim \begin{cases} \text{Gamma}(\mathbf{g}_k) & k = z_i \\ \text{Uniform}([0, d_0]) & k \neq z_i \end{cases} \quad (4.6)$$

where \mathbf{g}_k is the parameter set of the Gamma distribution, $\{\alpha_k, \beta_k\}$, for each cluster, and $\text{U}([0, d_0])$ is the uniform distribution function over $[0, d_0]$ with d_0 a large enough constant.

The goal is to estimate the unknown parameters, \mathbf{g}_k 's and w_k 's, and the membership likelihood of each trajectory to each cluster based on the values of the d_{ik} 's. Although in simple cases it is possible to infer the hidden parameters by finding the exact solutions of the Bayes net (defined later) using, for example, message-passing algorithms, exact variable elimination, or junction-tree algorithms, approximate methods are needed to deal with complex problems. The most common approaches are Monte Carlo and variational inference methods. In this chapter, we use the expectation-maximization algorithm to find the maximum likelihood (ML) solution in the absence of an atlas of white matter fiber tracts. In the next chapter, we adapt the EM approach for maximum a posteriori (MAP) estimation of parameters when an anatomical atlas is available.

4.4 Expectation-Maximization Algorithm

The Expectation-Maximization algorithm, first introduced by Dempster *et al.* [28], iteratively estimates the local maximum of the log likelihood of the data to find the maximum likelihood estimation of the unknown parameters:

$$\hat{\Theta} = \arg \max_{\Theta} L(\Theta), \quad (4.7)$$

where $L(\Theta) = \log p(D|\Theta)$, Θ is the collection of the unknown parameters, and $\hat{\Theta}$ is an estimate for Θ . This maximization is done by defining a lower bound on the incomplete log likelihood, $L(\Theta)$, as shown in Figure 4-5. Note that $p(D|\Theta)$ is called

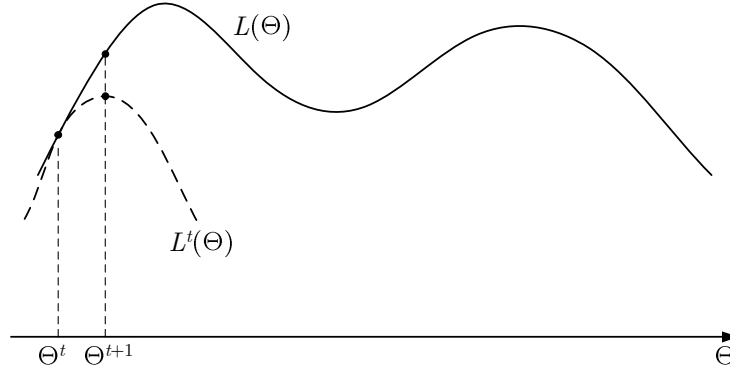


Figure 4-5: Schematic representation of procedure used by the EM algorithm to find the local maximum of the log likelihood function, $L(\Theta)$. In the expectation step the algorithm constructs a lower bound $L^t(\Theta)$ based on the current estimate of the parameter, Θ^t , of the optimal solution, $\hat{\Theta}$. In the maximization step, the maximum of the lower bound is calculated to obtain the updated parameter, Θ^{t+1} .

incomplete as it does not contain the hidden data, here the label of each trajectory, and similarly $p(D, Z|\Theta)$ is called the complete data likelihood.

Very generally speaking, the EM algorithm has two steps: in the expectation step (E-step), the algorithm defines a lower bound based on the current estimate, Θ^t , of the optimal solution $\hat{\Theta}$, while in the maximization step (M-step) the maximum of the lower bound is calculated to obtain the updated parameter, Θ^{t+1} . This procedure is repeated until convergence is achieved.

To construct $L(\Theta)$, $p(D|\Theta)$ is marginalized over the hidden data, Z :

$$p(D|\Theta) = \sum_Z p(D, Z|\Theta). \quad (4.8)$$

Now, the maximization problem takes the following form:

$$\hat{\Theta} = \arg \max_{\Theta} \log \left[\sum_Z p(D, Z|\Theta) \right]. \quad (4.9)$$

The essence of the EM algorithm [28] is that it re-defines the maximization problem in (4.7) by considering the following objective function, instead of using the data log

likelihood:

$$L_t(\Theta) \doteq E_{Z|D, \Theta^t} \left[\log \frac{p(D, Z|\Theta)}{p(Z|D, \Theta^t)} \right]. \quad (4.10)$$

It can be shown that this objective function has two main properties, as illustrated in Figure 4-5 [14]:

- It defines a lower bound on the incomplete log likelihood, i.e., $L_t(\Theta) \leq L(\Theta)$.
- It touches the incomplete log likelihood at the current estimate of the parameter, i.e., $L_t(\Theta^t) = L(\Theta^t)$ and $L'_t(\Theta^t) = L'(\Theta^t)$.

And with the above properties, it is proved that the estimate of the parameters approaches the local maximum of the incomplete log likelihood at each iteration [14].

The EM algorithm proceeds in two steps:

Expectation Step: Construct the lower bound $L_t(\cdot)$ of the incomplete log likelihood, $L(\cdot)$, based on the current estimate of the parameter, Θ^t :

$$L^t(\Theta) = E_{Z|D, \Theta^t} \left[\log \frac{p(D, Z|\Theta)}{p(Z|D, \Theta^t)} \right]. \quad (4.11)$$

Maximization Step: Update the parameter estimate by maximizing the current lower bound:

$$\Theta^{t+1} = \arg \max_{\Theta} L^t(\Theta). \quad (4.12)$$

From the two properties, stated above, it follows that the new estimate of the parameter is a better estimate of the optimal $\hat{\Theta}$:

$$L(\Theta^{t+1}) \geq L_t(\Theta^{t+1}) \geq L_t(\Theta^t) = L(\Theta^t). \quad (4.13)$$

This proves the convergence property of the EM algorithm if an upper bound on the log likelihood exists.

Generally the objective function of the EM algorithm is defined as:

$$L^t(\Theta) = E_{Z|D, \Theta^t} [\log p(D, Z|\Theta)] = E_{Z|D, \Theta^t} L_c(\Theta), \quad (4.14)$$

where $L_c(\Theta)$ is the complete log likelihood. This objective function has the same maximum as the objective function defined by (4.11), since the denominator $p(Z|D, \Theta^t)$ is constant with respect to Θ .

4.5 Implementation of Our Proposed Clustering Approach

The procedure of our proposed clustering method is shown in Algorithm 4.1. At the heart of this method is the EM algorithm to infer the membership probabilities and cluster parameters. Once the EM algorithm converges, the outer loop updates the cluster centers and re-computes the distance between trajectories and cluster centers. The outer loop is repeated until the cluster centers converge.

Details of calculating the distance between the trajectories and cluster centers (line 4 of the algorithm) were discussed in Chapter 3. Other steps are explained in this section. We start by presenting our formulation of the EM algorithm to solve the inference problem shown in the graphical model of Figure 4-4 (steps 6-14 of the algorithm).

4.5.1 Application of EM to clustering of white matter trajectories

We find the ML estimates of the parameter by maximizing the expectation of log likelihood of the complete data at each iteration as follows.

$$\Theta^{t+1} = \arg \max_{\Theta} E_{\mathbf{z}|\Theta^t, \mathbf{d}} [\log p(\mathbf{z}, \mathbf{d}|\Theta)]. \quad (4.15)$$

where $\Theta = \{\mathbf{g}, \mathbf{w}\}$, i.e., the collection of cluster parameters and mixing weights. We then assume that we observe N samples of the data, \mathbf{d} . It follows that:

Algorithm 4.1: The algorithm used to cluster fiber trajectories in the absence of an anatomical prior and to infer cluster parameters. At the heart of the algorithm are the steps of EM which updates the membership probabilities and cluster parameters. The outer loop re-calculates the cluster center once the EM step is converged.

Input: Set of 3D trajectories. $\{\mathbf{r}_i\}_{i=1}^N$ and initial centers $\{\boldsymbol{\mu}_k\}_{k=1}^K$
Output: Probabilistic assignment of trajectories to clusters

```

1 repeat
2   foreach trajectory  $\mathbf{r}_i$  do
3     | Compute the distances of the trajectory to all cluster centers  $\mathbf{d}_i$ ;
4   end
5   repeat                                     /* expectation-maximization */
6     | foreach trajectory  $\mathbf{r}_i$  do             /* E-step: */
7       | Calculate membership probabilities  $p_{ik}$ ;
8       | Remove outliers;
9     | end
10    | foreach cluster  $k$  do                     /* M-step */
11      | Update parameters of the cluster model  $\mathbf{g}_k$  and mixing weights  $w_k$ ;
12    | end
13  until convergence;
14  foreach cluster  $k$  do
15    | Update cluster center  $\boldsymbol{\mu}_k$ ;
16  end
17 until convergence ;

```

$$\begin{aligned}
\Theta^{t+1} &= \arg \max_{\Theta} \mathbb{E}_{\mathbf{z}|\mathbf{d}, \Theta^t} \left[\log \prod_{i=1}^N p(\mathbf{d}_i, \mathbf{z}|\Theta) \right] \\
&= \arg \max_{\Theta} \sum_{i=1}^N \mathbb{E}_{\mathbf{z}|\mathbf{d}, \Theta^t} [\log p(\mathbf{d}_i, \mathbf{z}|\Theta)] \\
&= \arg \max_{\Theta} \sum_{i=1}^N \sum_{k=1}^K p(z_i = k|\mathbf{d}_i, \Theta^t) \log [p(\mathbf{d}_i|z_i = k, \Theta)p(z_i = k|\Theta)] .
\end{aligned} \tag{4.16}$$

The last expression was obtained by expanding the expectation and applying Bayes rule on $p(\mathbf{d}_i, \mathbf{z}|\Theta)$. Note that $p_{ik} = p(z_i = k|\mathbf{d}_i, \Theta^t)$ is calculated in the E-step and is

independent of Θ . So we have:

$$\Theta^{t+1} = \arg \max_{\Theta} \sum_{i=1}^N \sum_{k=1}^K [p_{ik} \log p(\mathbf{d}_i | z_i = k, \Theta^t) + p_{ik} \log(p(z_i = k | \Theta))] . \quad (4.17)$$

The above expression is solved in two alternating steps:

Membership probabilities assignment(E-Step)

Assuming that the parameters of the clusters are known, using Bayes' rule, the probability that trajectory \mathbf{r}_i belongs to cluster k is

$$p_{ik} = \Pr(z_i = k | \mathbf{d}_i, \Theta^t) = \frac{\Pr(z_i = k) p(\mathbf{d}_i | z_i = k, \mathbf{g}_k)}{\sum_k \Pr(z_i = k) p(\mathbf{d}_i | z_i = k, \mathbf{g}_k)}, \quad (4.18)$$

where $p(\mathbf{d}_i | z_i = k, \mathbf{g}_k) = \text{Gamma}(d_{ik} | \alpha_k, \beta_k)$.

Updating model parameters(M-Step)

In this step, parameter estimates are updated. We first decompose Θ into its elements, \mathbf{g} and \mathbf{w} , to obtain the corresponding maximization expressions for each variable:

$$\begin{aligned} \mathbf{g}^{t+1} &= \arg \max_{\mathbf{g}} \sum_{i=1}^N \sum_{k=1}^K [p_{ik} \log p(\mathbf{d}_i | z_i = k, \mathbf{g}, \mathbf{w}) + p_{ik} \log(p(z_i = k | \mathbf{g}, \mathbf{w}))] \\ &= \arg \max_{\mathbf{g}} \sum_{i=1}^N \sum_{k=1}^K p_{ik} \log p(\mathbf{d}_i | z_i = k, \mathbf{g}) \end{aligned} \quad (4.19)$$

and

$$\begin{aligned} \mathbf{w}^{t+1} &= \arg \max_{\mathbf{w}} \sum_{i=1}^N \sum_{k=1}^K [p_{ik} \log p(\mathbf{d}_i | z_i = k, \mathbf{g}, \mathbf{w}) + p_{ik} \log p(z_i = k | \mathbf{g}, \mathbf{w})] \\ &= \arg \max_{\mathbf{w}} \sum_{i=1}^N \sum_{k=1}^K p_{ik} \log p(z_i = k) \end{aligned} \quad (4.20)$$

Note that z and \mathbf{g} are independent (unconditioned on the observed data, \mathbf{d}).

First let's solve the maximization equation for \mathbf{g} . Recall the likelihood function:

$$p(\mathbf{d}_i | z_i = k, \mathbf{g}) = \text{Gamma}(d_{ik}; \mathbf{g}_k) \prod_{j \neq k} U(d_{ij}; 0, d_0), \quad (4.21)$$

where $\mathbf{g}_k = \{\alpha_k, \beta_k\}$ is the collection of the parameters of the gamma distribution function for cluster k . The maximization equations are obtained by differentiating the following expression with respect to each parameter.

$$Q(\mathbf{g}) = \sum_{i=1}^N \sum_{k=1}^K p_{ik} \left(\log \text{Gamma}(d_{ik}; \mathbf{g}_k) + \log \prod_{j \neq k} U(d_{ij}; 0, d_0) \right) \quad (4.22)$$

The inverse scale parameter is obtained by $\partial Q(\mathbf{g}) / \partial \beta_k = 0$, which simply gives:

$$\sum_{i=1}^N p_{ik} \left(\frac{\alpha_k}{\beta_k} - d_{ik} \right) = 0 \quad (4.23)$$

or

$$\beta_k = \alpha_k \frac{\sum_{i=1}^N p_{ik}}{\sum_{i=1}^N p_{ik} d_{ik}}. \quad (4.24)$$

The shape parameter is given by $\partial Q(\mathbf{g}) / \partial \alpha_k = 0$, which gives:

$$\sum_{i=1}^N p_{ik} \left[\log d_{ik} + \log \alpha_k + 1 + \log \frac{\sum_{i=1}^N p_{ik}}{\sum_{i=1}^N p_{ik} d_{ik}} - d_{ik} \frac{\sum_{i=1}^N p_{ik}}{\sum_{i=1}^N p_{ik} d_{ik}} - \psi(\alpha_k) \right] = 0 \quad (4.25)$$

where $\psi(.) = \Gamma'(.)/\Gamma(.)$ is the digamma function. This gives:

$$\log \alpha_k - \psi(\alpha_k) = -\log \frac{\sum_{i=1}^N p_{ik}}{\sum_{i=1}^N p_{ik} d_{ik}} - \frac{\sum_{i=1}^N p_{ik} d_{ik}}{\sum_{i=1}^N p_{ik}}, \quad (4.26)$$

which does not have a closed-form solution. However, good approximations can be obtained by noting that

$$\log \alpha_k - \psi(\alpha_k) \approx \frac{1}{\alpha_k} \left(\frac{1}{2} + \frac{1}{12\alpha_k + 2} \right). \quad (4.27)$$

This gives:

$$\alpha_k \approx \frac{3 - x + \sqrt{(x - 3)^2 + 24x}}{12x}, \quad (4.28)$$

where

$$x = \log \left(\frac{\sum_i p_{ik} d_{ik}}{\sum_i p_{ik}} \right) - \frac{\sum_i p_{ik} \log(d_{ik})}{\sum_i p_{ik}}. \quad (4.29)$$

The maximization equation for the mixing weights, \mathbf{w} is given by:

$$\mathbf{w}^{t+1} = \arg \max_{\mathbf{w}} \sum_{i=1}^N \sum_{k=1}^K p_{ik} \log p(z_i = k) \quad (4.30)$$

with the additional constraint that $\sum_{k=1}^K p(z_i = k) = \sum_{k=1}^K w_k = 1$. To account for this constraint we add the Lagrange multiplier to the RHS of the above equation:

$$\mathbf{w}^{t+1} = \arg \max_{\mathbf{w}} \sum_{i=1}^N \sum_{k=1}^K p_{ik} \log w_k - \lambda \left(\sum_{k=1}^K w_k - 1 \right) \quad (4.31)$$

where λ is the Lagrange multiplier, and take the derivative to obtain:

$$w_k = \frac{\sum_{i=1}^N p_{ik}}{\lambda}. \quad (4.32)$$

The constraint $\sum_{k=1}^K w_k = 1$ gives

$$\lambda = \sum_{k=1}^K \sum_{i=1}^N p_{ik} = N. \quad (4.33)$$

So the updated mixing weights are calculated as

$$w_k = \frac{1}{N} \sum_{i=1}^N p_{ik}. \quad (4.34)$$

4.5.2 Updating the Cluster Centers and Distance Matrix

When the EM algorithm has converged, the membership probabilities, p_{ik} 's, are used to re-calculate the cluster centers, using the following seemingly simple expressions:

$$\boldsymbol{\mu}_k = \frac{\sum_i p_{ik} \mathbf{r}_i^{(k)}}{\sum_i p_{ik}} \quad (4.35)$$

$$\boldsymbol{\Lambda}_k = \frac{\sum_i p_{ik} (\mathbf{r}_i^{(k)} - \boldsymbol{\mu}_k)(\mathbf{r}_i^{(k)} - \boldsymbol{\mu}_k)^T}{\sum_i p_{ik}} \quad (4.36)$$

where $\mathbf{r}_i^{(k)}$ is the trajectory \mathbf{r}_i re-parameterized to have point correspondence to cluster k . In practice, we consider only trajectories that have a membership probability higher than a positive threshold (in our implementation this threshold is set to 0.01).

Note that in computing covariance matrices, it is possible for trajectories to have more than one point associated with a particular point on the cluster center. In that case, the closest one is considered. Also, note that ideally points that correspond to a particular point on the cluster center lie on the plane orthogonal to the curve at this point. This would be consistent with the fact the natural variation space of the points on the curve is only the orthogonal plane at each point. One could force the tangential component of the $\mathbf{r}_i^{(k)} - \boldsymbol{\mu}_k$ to zero by subtracting it, i.e.,

$$\boldsymbol{\Lambda}_k = \frac{\sum_i p_{ik} \boldsymbol{\Delta} \mathbf{r}_i^{(k)} \boldsymbol{\Delta} \mathbf{r}_i^{(k)T}}{\sum_i p_{ik}}, \quad (4.37)$$

with

$$\boldsymbol{\Delta} \mathbf{r}_i^{(k)} = (\mathbf{r}_i^{(k)} - \boldsymbol{\mu}_k) \left(1 - \frac{(\mathbf{r}_i^{(k)} - \boldsymbol{\mu}_k) \cdot \boldsymbol{\mu}_k}{\|(\mathbf{r}_i^{(k)} - \boldsymbol{\mu}_k) \cdot \boldsymbol{\mu}_k\|} \right). \quad (4.38)$$

The approach described above by itself does not allow the length of the initialized centers to evolve during the EM iterations. In fact, the number of points on the cluster center remains constant but the distance between successive points changes. We desire, however, that the method be less dependent on the initial centers and flexible enough so that the centers of clusters are able to grow or shrink in length. To achieve this goal we perform a uniform re-sampling of the updated centers after the M-step at

each iteration. In doing so, first the spline representation of centers is reconstructed from their updated samples, and then the arc length of the reconstructed center is uniformly sampled. This also ensures that cluster centers remain smooth.

4.5.3 Initialization

When starting the iterative EM algorithm with the E-step, all the parameters, Θ , must be initialized. We set the shape parameter of the Gamma distribution equal to one, i.e., $\alpha = 1$, to have an exponential distribution which favors those trajectories most similar to the initial center. The algorithm is not very sensitive to the initial value of the inverse scale parameter, β , initialized to ten in our implementation. However, one should be careful not to set it too small, otherwise many trajectories are labeled as unclustered in the initial step. On the other hand, if β is set too large, the likelihood would be very small even for trajectories very similar to the cluster center. Figure 4-6 illustrates the evolution of the Gamma distribution for the clusters of the first case shown in Figure 5-12. Convergence is achieved after just a few iterations of the EM algorithm. As stated earlier, the Gamma distribution is initialized with $\alpha = 1$, corresponding to a exponential distribution, to value those trajectories that have no distance to the initial cluster center. However, as the algorithm proceeds, the Gamma distribution evolves from a very broad distribution to a narrow distribution with small but non-zero mode.

In our implementation, we have a minimum threshold on the membership likelihood to ensure the robustness of the algorithm, as discussed in the next subsection. The cluster centers can be initialized manually by selecting a set of trajectories with particular shapes from the data. As an alternative, the initial cluster centers can be supplied by an atlas of fiber tracts, if available, such that the mean trajectories of atlas clusters are employed after registration of the case and the atlas. Dependence of the algorithm on the initial centers will be discussed in Section 4.6. The associated covariance matrix of each center initially does not play a significant role so we set it to the unity matrix. This reduces the computation of the Mahalanobis distance to Euclidean distance through the similarity measurement, which is an acceptable

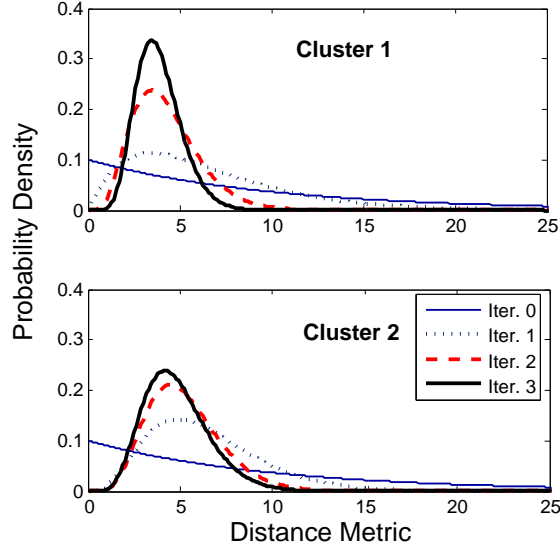


Figure 4-6: Evolution of the Gamma distributions that characterizes the normalized distance metric for the two clusters shown on the left in Figure 5-12. After just a few iterations, the distribution converges to a narrow distribution with small but non-zero mode.

approximation for the first iteration of the algorithm.

4.5.4 Handling Outliers

In mixture-model clustering, it is assumed that each data point is modeled by the mixture of a finite number of component densities. However, in our case, there might be trajectories resulting from the tractography which do not resemble any of the user- or atlas-initialized cluster centers or are simply generated by noise and imperfections in the data. These outliers will lead to increased variance of the density functions if not properly handled. This would result in excessive spread of the bundles or even instability of the algorithm. An outlier is identified by imposing a threshold on the membership likelihoods. If the membership likelihood of a given trajectory in *all* clusters is less than the specified threshold, that trajectory will be removed from further data processing. In fact, with this threshold the heterogeneity within each cluster is controlled. The larger the threshold is, the more compact are the resulting bundles, and consequently the greater is the number of unclustered trajectories. Handling outliers is not straightforward in previously proposed clustering schemes [17, 56, 64].

Unlike those methods, we allow the distribution of each cluster to have a different set of parameters (α_k 's, β_k 's), all inferred from the data by using the EM algorithm, and hence the user needs to set only the mentioned threshold to effectively remove the outliers.

4.6 Results

In this section we illustrate different aspects of our proposed approach, such as its features for performing population studies, and its performance on probabilistic assignments of trajectories to bundles and outlier rejection. We also discuss the sensitivity of the output to the specified initial centers. We leave the results on quantitative analysis and visualization of the fiber bundles to Chapter 6.

4.6.1 Impact of Initial Cluster Centers

To investigate the sensitivity of the clustering to the initial centers, we randomly selected different sets of trajectories from each cluster shown in Figure 4-7(a). At each run with one of the sets as the initial centers as shown in part (b), the final centers obtained by the clustering algorithm are almost identical as shown in Figure 4-7(c). This demonstrates the robustness of the algorithm with respect to the variations in the initial centers within each cluster.

Note that although the sensitivity to cluster centers is not desired in the sense that results depend on which trajectory is selected within the cluster, it is obvious that the algorithm must be responsive to the selected initial centers as the representatives of the expected clusters. Figure 4-8 shows a real clinical application in which the clustering of trajectories into three bundles of Inferior Longitudinal Fasciculus (ILF), Inferior Fronto-Occipital Fasciculus (IFO) and Uncinate Fasciculus (UF) is desired. As shown in this figure, our method successfully clusters the spatially adjacent trajectories into the mentioned bundles by selecting three initial centers, each representing one of the expected clusters.

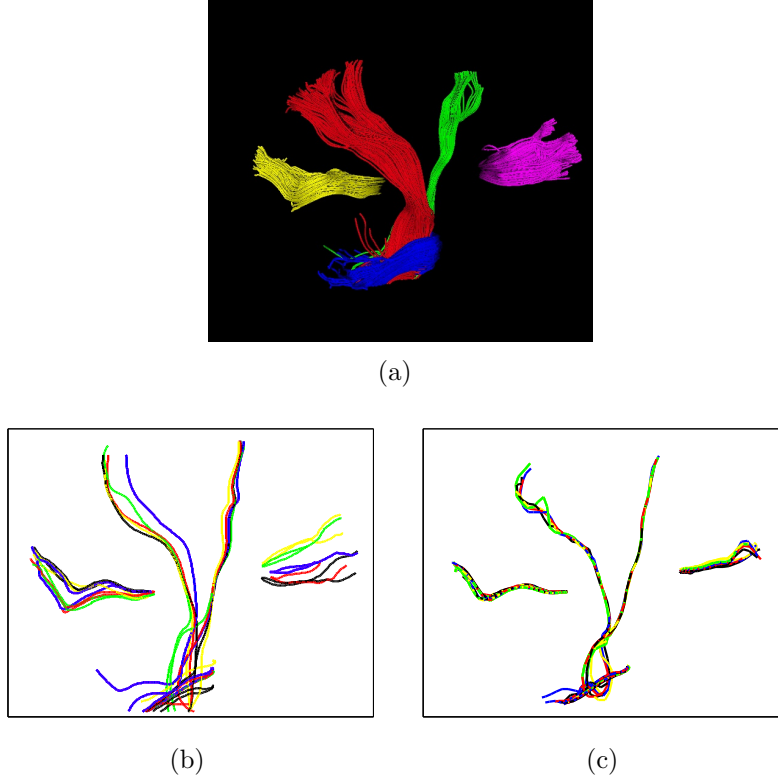


Figure 4-7: (a) Trajectories of 5 different clusters used for quantitative analysis: splenium (yellow), corticospinal (red), corticobulbar (green), middle cerebellar peduncle (blue), and genu (magenta). Robustness of the EM algorithm with respect to the initial cluster centers is clearly seen. The algorithm was run 5 times with different initial centers (b) to cluster the trajectories in (a). Final cluster centers collected in (c) show little dependence on initial centers.

4.6.2 Outlier Rejection

As stated earlier in Section 4.5.4, outliers are identified by imposing a threshold on the likelihood functions in the EM algorithm². We define this threshold as a ratio ranging between 0 and 1 in terms of the peak of the likelihood functions. Figure 4-9 shows the experimental results of clustering superior longitudinal fasciculus (SLF) to arcuate fasciculus (the curved bundle) and the upper part of SLF, by manually specifying two trajectories, one from each of the expected clusters, as the initial cluster centers. Part (a) of the figure shows the unclustered trajectories, while part (b) shows the trajectories colored based on their membership probability to each of

²Note that the outlier rejection cannot be done by thresholding the posterior probabilities as those are being normalized.

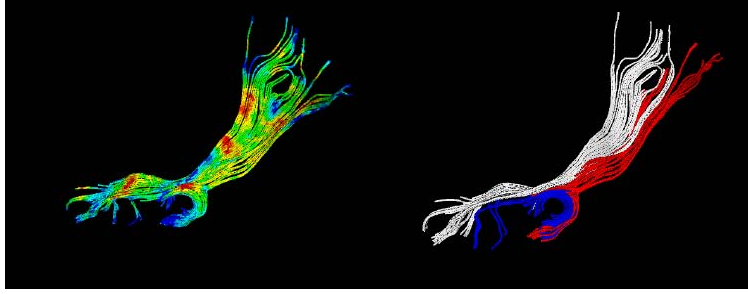


Figure 4-8: Clustering of the trajectories (colored based on the local FA value) shown in (a) to Inferior Longitudinal Fasciculus (white), Inferior Fronto-Occipital Fasciculus (red) and Uncinate Fasciculus (blue) clusters shown in (b).

the two clusters. Initial cluster centers are overlaid on this plot in white, while final cluster centers are shown in yellow. Note that although the initial centers are not the best possible choice, the algorithm has converged to the medial part of the clusters. Trajectories are identified by setting a threshold of 0.2 on the likelihood function and are shown in part (c). Most of the outliers identified by the algorithm are actually part of another bundle called the short arcuate fibers [55].

Figure 4-10 shows how the spread and homogeneity of the fiber bundles is controlled by the likelihood threshold. As expected, with the threshold of 0.6, shown in part (b) of this figure, more compact bundles are achieved, compared to those obtained with a threshold of 0.2, shown in (a). The choice of the threshold depends on our knowledge from anatomy and the application. As will be discussed in the next chapter, the use of an anatomical atlas specifies limits on the spatial spread of the bundles and so makes the algorithm more robust with respect to specified parameters.

4.6.3 Whole Brain Analysis and Population Studies

The computation of the similarity measure is usually the most computationally expensive step in clustering algorithms used for fiber trajectories. As discussed in the previous chapter, our approach which uses the distance map for similarity measure calculation makes the computation time increase slightly with respect to an increase in number of trajectories. This feature makes the analysis of a large number of trajectories, which is encountered in whole brain or population studies, possible with the

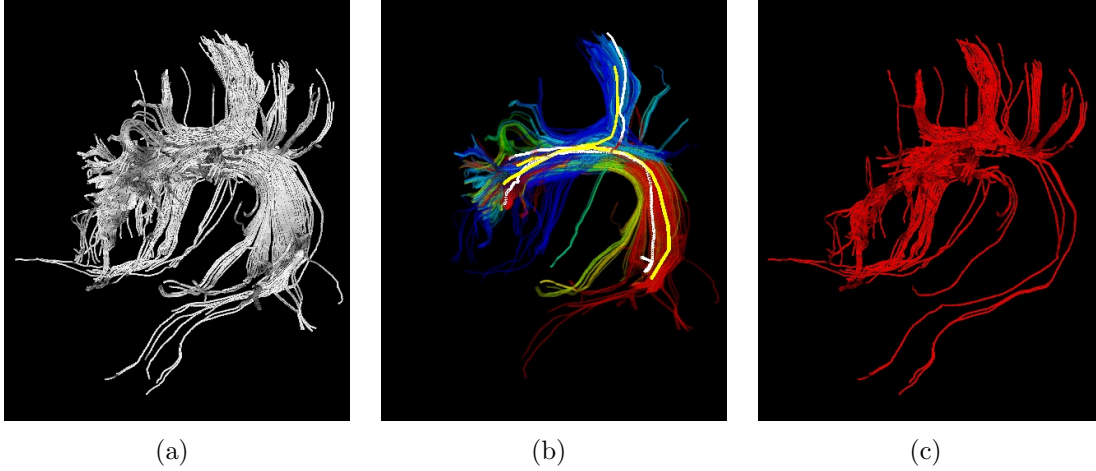


Figure 4-9: Demonstration of outlier rejection by imposing a threshold on the likelihood function. (a) Unclustered trajectories from superior longitudinal fasciculus. (b) Probabilistic assignment of these trajectories into arcuate fasciculus and the upper portion of SLF. Initial cluster centers are shown in white, while final cluster centers are shown in yellow. (c) Trajectories identified as outliers with a threshold of 0.2 on the likelihood function.

same level of accuracy as for analysis of small data. Unlike other methods proposed for clustering of fiber trajectories, no dimensionality/size reduction is needed.

Figure 4-11 shows the clustering results for roughly 3000 trajectories from corpus-callosum, middle cerebellar peduncle, corticobulbar, and corticospinal tracts into 25 bundles. As the initial centers, 25 trajectories from the data were selected manually, each representing an expected cluster³. With a threshold equal to 0.2, 45 trajectories remained unlabeled. Note that the trajectories in Figure 4-11 are colored based on their maximum membership probabilities.

A more interesting example is clustering of a single bundle of interest in a population of subjects. In this experiment we cluster the cingulum bundle, which is one of the difficult bundles of fiber tracts to cluster. Even when seeded from a user-defined ROI, the tractography step leads to a set of disordered trajectories, mostly short in length because of low fractional anisotropy. Also, due to the adjacency to the corpus callosum, many callosal trajectories are included, which adversely affects any further analysis of the bundle. Figure 4-12 shows the ROIs corresponding to different fiber

³As will be discussed in the next chapter, the initial centers can also be specified by the use of an anatomical atlas of fiber tracts.



Figure 4-10: Control of the spread of the bundles with the likelihood threshold. With a threshold of 0.6 more compact trajectories are obtained (b) compared to those obtained with a threshold of 0.2 (a).

bundles. Cingulum bundles are the two narrow C-shaped structures on the top of the U-shaped corpus collsum, colored in purple and light green.

The ROI is specified by an expert and streamline tractography is performed on each case. All of roughly 5000 trajectories are then mapped into the same space by applying the transformation matrix which has been calculated by registering the FA volumes of the subjects by an affine transformation(See figure 4-13). We discuss the registration process and the sensitivity of the method to errors in registration in the next chapter. As shown in Figure 4-14 for eight subjects, our method is well capable of clustering these trajectories into the desired bundles. Two arbitrary trajectories, one from the superior and one from the posterior part of the cingulum were selected as the initial cluster centers. Knowledge of the point correspondence and rigorous calculation of the similarity measure is essential for clustering of such a disordered set of trajectories. This example also shows the capability of the method in automating the clustering on a population of subjects.

In terms of computational complexity, our approach is superior to computing the bidirectional Hausdorff distance. The computation time of our algorithm is approximately equal to: $I \times K \times (N \times T_t + T_c + T_{dm})$, where I is the number of EM iterations, T_t is the processing time each trajectory requires to calculate its distance from a

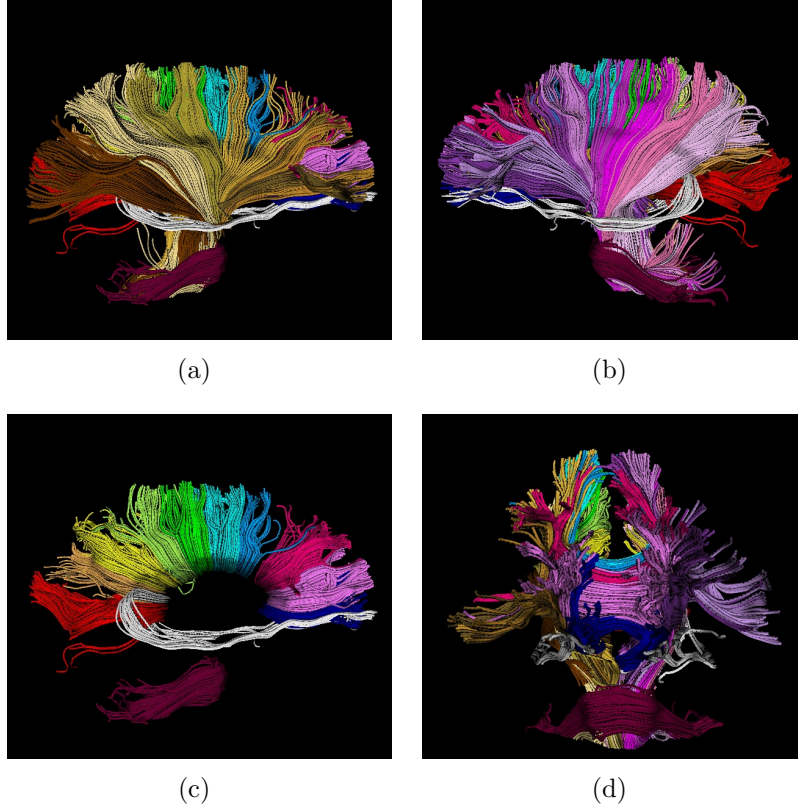


Figure 4-11: About 3000 trajectories clustered into 25 user-initialized bundles. Clusters include different segments of the corpus callosum, tapetum, middle cerebellar peduncle, corticobulbar and corticospinal tracts, and different portions of thalamic radiation.

cluster center and the associated membership likelihood, T_c is the processing time to update the cluster parameters, which grows linearly with N and T_{dm} is the computation time of constructing a distance map from a given cluster center. Note that there exist fast methods to compute the Euclidean distance map in linear time with respect to the number of voxels in the space that the distance map is constructed in. The computational complexity in our algorithm increases linearly with the number of trajectories, whereas performing a common curve matching technique or Hausdorff distance is of the order of N^2 as it would require pair-wise calculations. In terms of memory consumption, although one distance map is computed for each cluster in our approach, only one distance map is stored in the memory at a time.

To demonstrate the time efficiency of our method in handling a large number of trajectories, we record the CPU time for clustering of around 21000 trajectories

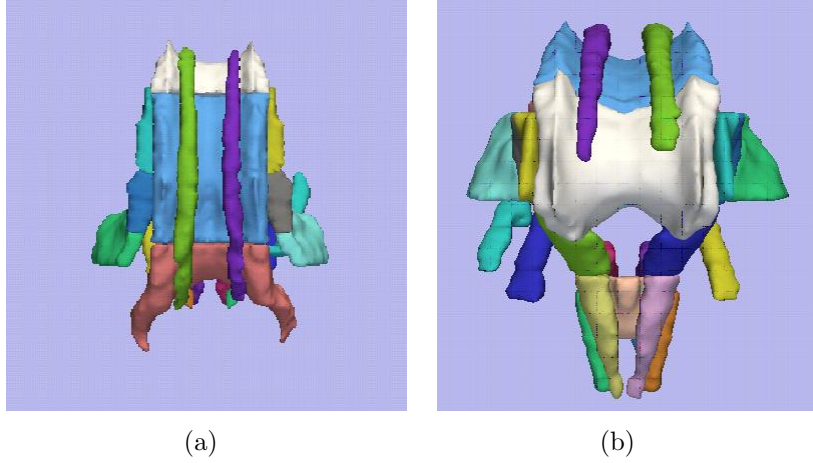


Figure 4-12: Cingulum bundles are the two narrow C-shaped structures on the top of the U-shaped corpus collsum, colored in purple and light green. Due to the adjacency to the corpus callosum, many callosal trajectories are included in tractography of these bundles that adversely affect any further analysis on them.

extracted for the left Uncinate Fasciculus in 44 subjects. This experiment is repeated for a downsampled version of such data and the recorded processing times are shown in Figure 4-15 in a logarithmic scale. All of the runs in this experiment are performed with the outlier threshold of 0.2 and three iterations of EM. The recorded processing times are shown with open squares as a function of the number of trajectories. For a small number of trajectories, there is a fixed processing time, which is in fact equal to the computation time of the distance map, roughly 10 seconds per iteration. The filled triangles in the figure are the processing times minus 30 seconds required for computing the distance map (three EM iterations multiplied by 10 seconds). These data points lie on a line with a slope of approximately one in the log-log scale, which demonstrates that the processing time grows linearly with the number of trajectories. This experiment confirms the formulation for processing time, presented in Section 3.4.

4.7 Summary

In this chapter we developed an expectation maximization algorithm to cluster fiber trajectories in a gamma mixture model context. The gamma distribution enabled

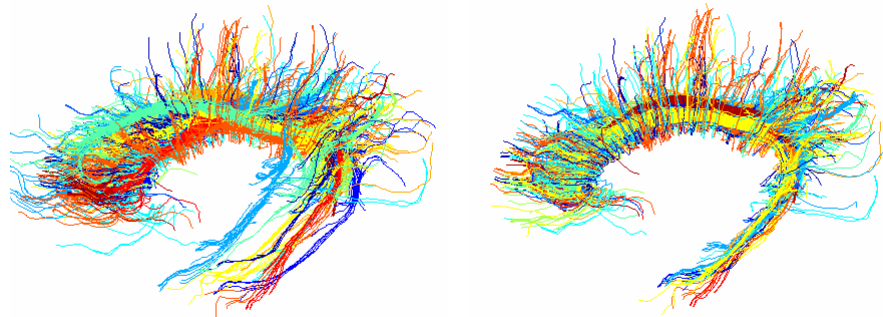


Figure 4-13: Registration process performed on cingulum trajectories from 8 different cases as a part of the pre-processing from clustering. Sagittal view of the trajectories before (left) and after (right) registration.

us to effectively model the normalized distance of the trajectories from each cluster center. A method was proposed to remove outlier trajectories by imposing a threshold on the membership likelihoods. Experimental results demonstrate the effectiveness of the proposed clustering method to group fiber trajectories into bundles based on their shape similarity and spatial proximity. Some anatomical information is included in the clustering in the form of the initial cluster centers. Hence unlike unsupervised clustering methods, the clusters correspond to actual anatomical bundles. The next chapter extends the EM algorithm developed in this chapter to include further prior knowledge to guide the clustering algorithm.

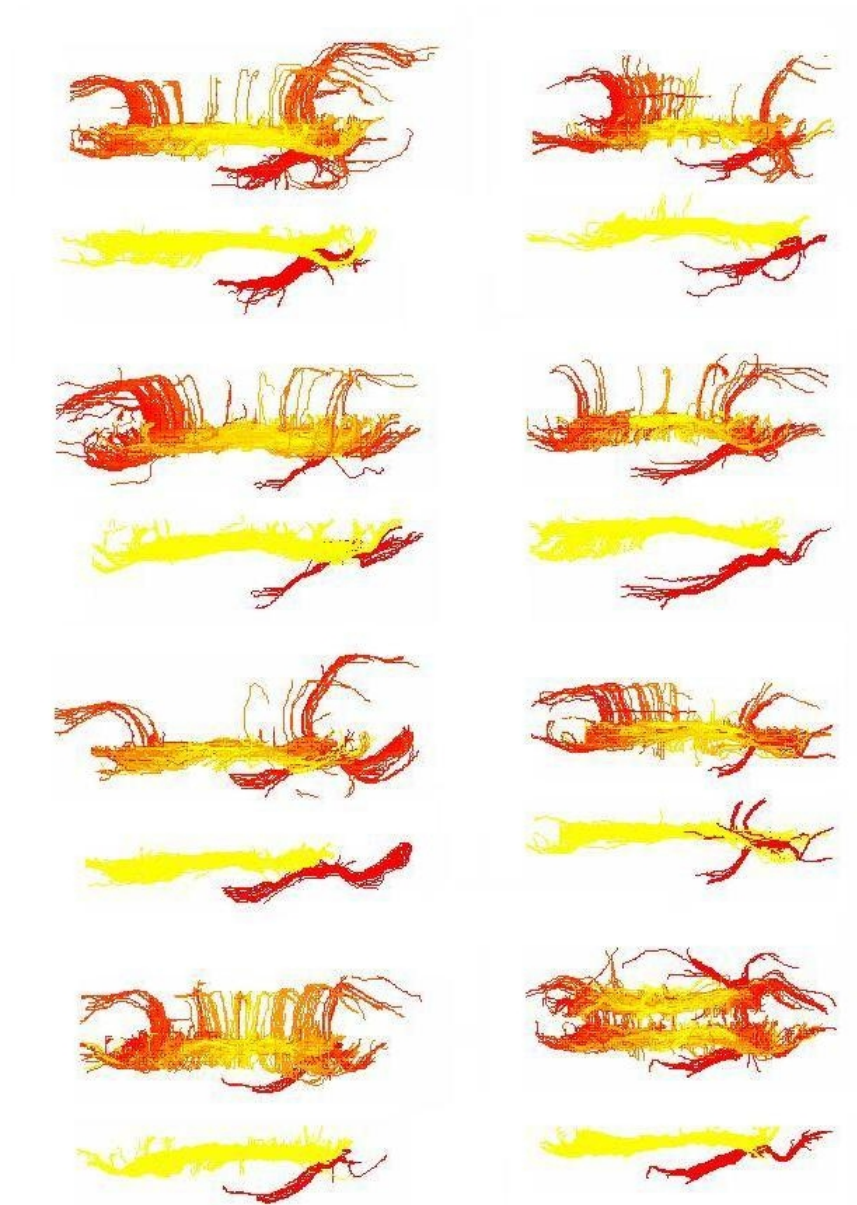


Figure 4-14: Clustering performed on cingulum trajectories from 8 individual cases. For each case the axial view of the trajectories before (top) and after (bottom) clustering is shown. Trajectories are clustered into superior and anterior part of the cingulum, shown in yellow and red, respectively. Note that the clustered bundles in each subject have a similar thickness with a single specified threshold for the population of trajectories.

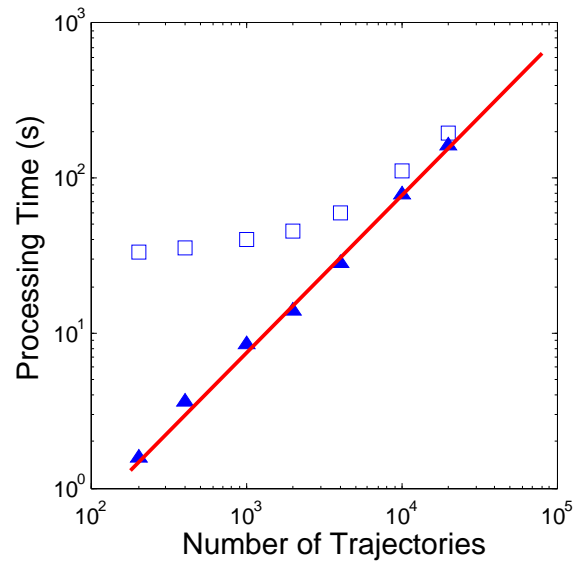


Figure 4-15: Processing time as a function of the number of trajectories measured for clustering of the left uncinute fasciculus from 44 cases. Open squares show the whole processing time, where its minimum is limited by the computation of the distance maps. Filled triangles show the processing time minus the time required to compute the distance map, demonstrating a linear dependence on the number of trajectories, i.e., a slope of approximately one in a log-log scale.

Chapter 5

Integration of Anatomical Information into Clustering of White Matter Trajectories

In this chapter we extend the clustering approach presented in the previous chapter to include spatial anatomical information from an atlas of fiber tracts. We believe that atlas-guided clustering that benefits from anatomical information not only produces anatomically meaningful clusters, but also yields more robust results that are less sensitive to the presence of outliers and imperfections in the DT-MRI data.

We propose two levels for integration of anatomical knowledge. In the first one, the spatial priors are a set of fixed parameters taken from the atlas, whereas in the second one, they are assumed to be random variables drawn from a Dirichlet distribution. In the first implementation with the fixed-weight prior (FWP), the anatomical atlas is imposed strongly on the clustering algorithm, whereas in the second approach with the adjustable-weight prior (AWP), the influence of the atlas can be controlled through a parameter. That weighting parameter can be set manually based on the expected compatibility of the subject with the normal population whose data was used to build the atlas. For instance, if the analysis is performed on a pathological case where deviation of fiber bundles from normal brain anatomy is expected, the weighting parameter needs to be set to the lowest possible. In the case that multiple

atlases are available, as will be discussed in this chapter, the weighting parameter can be set automatically based on the correlation between the atlases.

The atlas is not only used to guide the clustering algorithm here, but it also provides the seed points for the tractography and the initial settings of the EM algorithm. The proposed approaches provide a robust and automated tool for tract-oriented analysis, both in a single subject and over a population.

In this chapter, we first describe the possibilities for encoding the anatomical information available for fiber bundles. Then we review earlier attempts to incorporate anatomical knowledge, and finally present our formulation of the EM algorithm for the two approaches mentioned above.

5.1 Representation of Anatomical Knowledge

Atlases are the common way of representing anatomical knowledge in MRI studies. Such atlases are usually constructed by labeling regions of interest by an expert (or semi-automatically by an algorithm) on a large number of subjects, followed by a spatial normalization. And such an atlas is then used to label the same regions in a novel subject. As an example of an atlas, in segmentation of the brain to white matter, gray matter and CSF, each voxel of the atlas is associated with three numbers, which represent the membership probability of that voxel to each of the three tissues.

Recently, atlas construction has attracted significant attention in DT-MRI analysis. The level of representation is more varied in DTI compared to conventional MRI as the orientation of the fiber bundles also can be encoded in addition to their spatial position in the brain.

One way of representing an atlas of fiber tracts is to perform tractography on a set of healthy subjects, map the trajectories into a common space, and manually label them to represent their association to anatomical bundles [58]. Figure 5-1 illustrates such an atlas in which each trajectory is colored based on the anatomical bundle to which it belongs as defined by an expert. In an automatic atlas construction approach presented in [65], the trajectories from multiple subjects are clustered in a

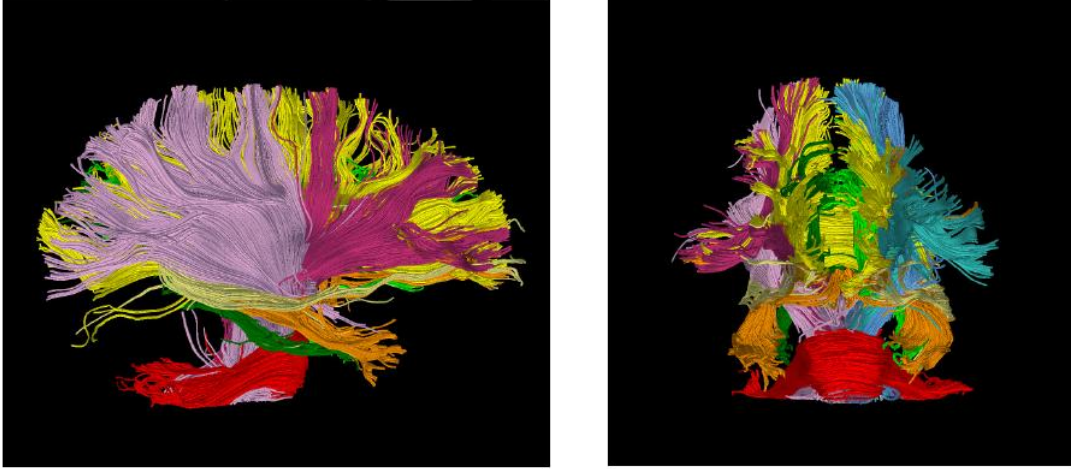


Figure 5-1: The constructed atlas of fiber tracts shown for the obtained bundles in the sagittal (a) and anterior coronal (b) view [58].

high-dimensional embedding space where each trajectory is represented by a point, and the cluster centroids are labeled by an expert to serve as a high-dimensional atlas. A similar atlas is constructed by simultaneously optimizing the parameters of registration and clustering [101]. Multiple atlases of diffusion images of the human brain have been constructed by Mori *et al.* at John Hopkins University. In the ICBM-DTI-81 white-matter labels atlas, 50 white matter tract labels were created by hand segmentation of a standard-space average of diffusion MRI tensor maps from 81 subjects with mean age of 39. In another JHU white-matter tractography atlas, 20 structures were identified probabilistically by averaging the results of running deterministic tractography on 28 normal subjects with a mean age of 29.

Regardless of the method used to construct and represent the atlas, one can define a rule to obtain a cluster membership probability for each trajectory, once the subject is mapped into the atlas space, to represent its membership in each anatomical fiber bundle. The methods that we present in this chapter are general in the sense that they only require a probability vector for each trajectory that is obtained from the atlas.

5.2 Related Work

Recently, employing anatomical atlases of fiber tracts in DT-MRI processing and analysis has attracted significant attention. Examples are atlas-guided tractography [62] and clustering or classification of the fiber trajectories [58].

We believe that the first attempt to use anatomical information goes back to our earlier work [58]. In that work, the atlas was represented by a set of labeled trajectories in a reference subject. The trajectories from a given subject were classified based on the pairwise distance between their B-spline coefficients and those of all trajectories from the atlas. This was done after spatial normalization of the baseline images by applying the obtained affine transformation to map the trajectories of the subject into the atlas space. The method is able to cluster the fiber tracts into anatomically known bundles. The correspondence of clusters across different subjects is also defined by default. Hence, unlike most clustering methods previously proposed, no post-processing is required to determine the correspondence between the bundles. However, this method relies solely on the atlas and ignores the variability among the trajectories in a bundle, which makes the result very sensitive to the atlas.

In another work, a high dimensional atlas is constructed in [65] in an embedding space. The atlas does not play any role in the clustering process and is used to associate groups of fiber trajectories to anatomical structures after they are clustered and mapped into an embedded space.

More recently, an atlas-based quantitative analysis of white matter fiber tracts has been proposed in which a probabilistic parcellation map of the tracts is used to obtain the weighted average of a set of parameters [41]. This approach relies exclusively on the atlas data and hence does not take into account the coherence among the trajectories in each bundle. Furthermore, the quantitative analysis is limited to averages over the entire ROI or slices perpendicular to the main axes of the atlas coordinate system since point correspondences between the trajectories are not calculated.

In the previous chapter, we proposed a clustering framework in which the anatom-

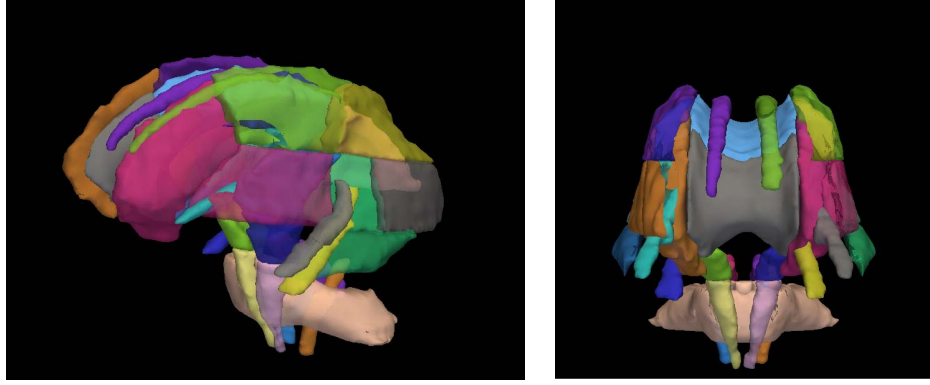


Figure 5-2: Visualization of some of the ROIs outlined by the atlas. These ROIs correspond to the major anatomical fiber tracts.

ical knowledge is included through specifying a number of prototype trajectories as the initial cluster centers. In this chapter, we extend our method to rigorously incorporate *spatial* anatomical information from an atlas of fiber tracts [92] in clustering of trajectories.

5.3 Membership Assignments from the Atlas

In this thesis, we use the atlas created by Mori *et al.* (<http://lbam.med.jhmi.edu>), which consists of 50 labeled regions that correspond to major anatomical bundles of fiber tracts in the human brain. As mentioned earlier, this atlas has been constructed by hand segmentation of a standard-space average of diffusion MRI tensor maps from 81 subjects with mean age of 39. This data is represented in MNI (Montreal Neurological Institute) space with $181 \times 217 \times 181$ mm³ isotropic resolution. Visualization of some of these regions in the 3D Slicer (www.slicer.org) is shown in Figure 5-2. To allow a probabilistic assignment at the region boundaries, we apply a Gaussian kernel with standard deviation of 2 within a $5 \times 5 \times 5$ window to each region.

Trajectories are extracted for each subject using a streamline tractography method [10] (see Algorithm 1 in Chapter 2) and mapped into the MNI atlas space. Note that the streamline tractography needs seed point specification, which is usually done by user interaction, to specify an ROI for each case. We use the ROIs from the atlas to automate this process. We first map the ROIs from the atlas into each subject’s space

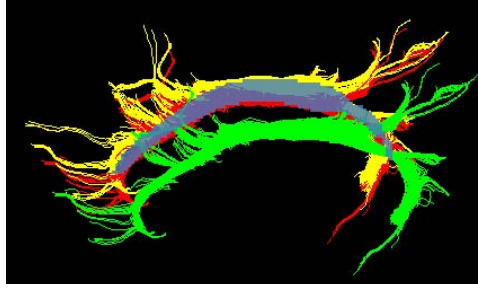


Figure 5-3: Demonstration of the registration process for the trajectories of superior cingulum in one of the subjects. Original trajectories (green) and trajectories mapped into the atlas space with either an affine registration (red) or a congealing registration to a common space for all subjects and then by an affine registration into the atlas space (yellow) are shown. The spatial extent of the cingulum specified by the atlas is also shown for comparison.

and perform a morphological dilation to ensure that the whole bundle is captured. The size of the structure element to perform the dilation operation is set to 3 for small bundles such as uncinate fasciculus and 1 for large bundles such as the splenium part of the corpus callosum.

Registration is performed on the corresponding maps of the fractional anisotropy (FA) ¹ to normalize for brain geometry, and then the obtained transformation is applied to the trajectories. An affine registration based on mutual information (MI) [95] usually gives satisfactory results as reflected in Figure 5-3. However, for population studies we opted to first map the subjects into a common space using the congealing algorithm [103] followed by affine registration to the MNI space. We decided to use this approach, as opposed to a series of pair-wise subject-template registrations, in order to avoid introducing bias in the population analysis. Figure 5-3 shows the results of registering the trajectories from the superior cingulum to the atlas space for one of the subjects.

With the trajectories projected to the atlas space, the membership probability for each trajectory, π_{ik} , is calculated by summing up the probabilities of its overlapping voxels with the probability maps of the fiber tracts in the atlas, and normalizing with

¹Registering the FA map does not produce the most accurate mapping as it does not use the whole tensor information. We made the choice of registering an scalar map since robust algorithms were available for that purpose and among the scalar maps, we chose FA as it is the most popular parameter of interest in quantitative clinical studies.

the volume of each tract in the atlas. The membership probabilities of each trajectory are then normalized, so that $\forall i, \sum_{j=1}^K \pi_{ij} = 1$. Note that with this implementation, the atlas ROIs provide the spatial prior, while the information about the shape and orientation of the tracts are captured by a representative curve for each bundle, used as the initial cluster center. Unlike a voxel-based method in which individual voxels (of each trajectory) receive their own membership probability [41], in our approach the probability is assigned to the entire trajectory and hence the method is less sensitive to local errors in registration.

5.4 Fixed-Weight Prior (FWP)

In this section, we propose a Bayesian approach for incorporating spatial anatomical information in the clustering of fiber trajectories. The essence of the approach is similar to the method presented in the previous chapter. The difference is that here we have access to prior information on the membership of each trajectory relative to each cluster. We use the EM formulation to estimate this information probabilistically.

5.4.1 EM Formulation for FWP

As mentioned earlier, once the trajectories are extracted from the DT data, they are mapped into the atlas coordinate system. Each trajectory \mathbf{r}_i then takes a membership probability $\boldsymbol{\pi}_i = [\pi_{i1}, \dots, \pi_{iK}]$, where each π_{ik} element denotes the atlas-specified membership of \mathbf{r}_i relative to cluster k . Note that the vector $\boldsymbol{\pi}_i$ is fixed throughout the iterations of the clustering algorithm. The number of clusters, K , is a subset of anatomical bundles in the atlas. Initial cluster centers are provided by the atlas and the distance vectors are computed as described in Chapter 3. In this setup, we have two sources of information for inferring the labels of trajectories, one from the atlas and one from the distance values to the cluster centers. The corresponding graphical model is depicted in Figure 5-4. The goal is again to infer unknown parameters, Θ ,

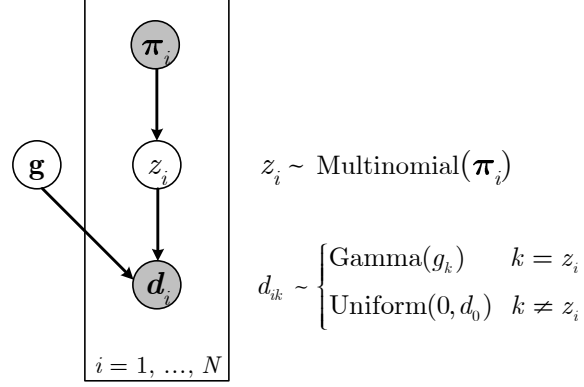


Figure 5-4: A Bayesian network that describes the dependencies among different variables. $\boldsymbol{\pi}$: atlas prior, z : trajectory label, \mathbf{d} : distance between trajectory and cluster centers and \mathbf{g} : parameters of the gamma distributions. Shaded nodes represent the observed data.

where here $\Theta = \mathbf{g}$, i.e., only the collection of parameters of the gamma distributions:

$$\Theta = \arg \max_{\theta} p(D, Z | \Theta), \quad (5.1)$$

where D denotes the distance values and Z represent the labels.

We employ Expectation-Maximization for the estimation of the parameters:

Expectation Step

In the expectation step, given the current estimates of the parameters, $\boldsymbol{\Theta}^t$, the membership probabilities are calculated.

$$p_{ik} = \Pr(z_i = k | \mathbf{d}_i, \boldsymbol{\Theta}^t) = \frac{\Pr(\mathbf{d}_i | z_i = k, \boldsymbol{\Theta}^t) \Pr(z_i = k | \boldsymbol{\Theta}^t)}{\sum_j \Pr(\mathbf{d}_i | z_i = j, \boldsymbol{\Theta}^t) \Pr(z_i = j | \boldsymbol{\Theta}^t)}.$$

Assuming that Z and $\boldsymbol{\Theta}^t$ are independent, the class assignment is independent of the model parameters:

$$\Pr(z_i = k | \boldsymbol{\Theta}^t) = \Pr(z_i = k) = \pi_{ik},$$

which is supplied by the atlas (see the previous section). So,

$$p_{ik} = \Pr(z_i = k | \mathbf{d}_i, \boldsymbol{\Theta}^t) = \frac{\text{Gamma}(d_{ik}; \Theta_j^t) \pi_{ik}}{\sum_{j=1}^K \text{Gamma}(d_{ij}; \Theta_j^t) \pi_{ij}}. \quad (5.2)$$

Maximization Step

The new estimates of the model parameters are updated in the maximization step:

$$\begin{aligned}\Theta^{t+1} &= \arg \max_{\Theta} (E_{Z|D, \Theta^t} [\log p(D, Z|\Theta)]) \\ &= \arg \max_{\Theta} (E_{Z|D, \Theta^t} [\log p(D|Z, \Theta) + \log p(Z|\Theta)]) .\end{aligned}$$

Since we assume Z to be independent of Θ , the last term can be eliminated from the maximization. Also, the \mathbf{d}_i 's are independent and each \mathbf{d}_i only depends on z_i . Hence,

$$\begin{aligned}\Theta^{t+1} &= \arg \max_{\Theta} \sum_{i=1}^N E_{Z|D, \Theta^t} [\log p(\mathbf{d}_i|\Theta, z_i)] \\ &= \arg \max_{\Theta} \sum_{i=1}^N \sum_{j=1}^J \Pr(z_i = j|d_i, \Theta^t) \log p(\mathbf{d}_i|\Theta, z_i = j),\end{aligned}\tag{5.3}$$

where $\Pr(z_i = k|d_i, \Theta^t)$ is given by (5.2), calculated in the expectation step. The maximization expression is similar to what we previously derived in the absence of the atlas prior in Chapter 4, as the prior only appears in the expectation step. So, to update the gamma parameters we have:

$$\alpha_k \approx \frac{3 - x_j + \sqrt{(x_j - 3)^2 + 24x_j}}{12x_j},\tag{5.4}$$

where

$$x_k = \log \left(\frac{\sum_i p_{ik} d_{ik}}{\sum_i p_{ik}} \right) - \frac{\sum_i p_{ik} \log(d_{ik})}{\sum_i p_{ik}},\tag{5.5}$$

and

$$\beta_k = \alpha_k \sum_i p_{ik} / \sum_i p_{ik} d_{ik}.\tag{5.6}$$

Once the EM algorithm converges, we update the cluster centers and recompute the distance vectors. The outliers are identified in the expectation step. If the membership likelihoods of a trajectory in all clusters are less than a user-specified threshold, that trajectory is identified as an outlier and is removed from further data processing.

Figure 5-5 compares the clustering results obtained with and without incorporating the atlas prior for the superior cingulum. The trajectories are colored based on their membership probability in each case. To emphasize the effect of the atlas, a worst-case scenario is presented in which the parameters that control the extent of the clusters are set such that the algorithm only excludes those trajectories that receive very small membership probability. Without the atlas, the algorithm gives moderate membership probability to those trajectories that are not very close to the initial center. However, as the algorithm proceeds, the cluster center drifts as shown in Figure 5-5 (a), so that these trajectories receive higher and higher membership probabilities. Even though the clustering results might still be acceptable, the cluster center is deformed, introducing significant error in the quantitative analysis. If the parameters are set in the correct range, the extent of the cluster can be controlled in order to prevent the inclusion of outliers and excessive drift of the cluster center. A trade-off should be made between the homogeneity of the clustered trajectories and the number of unclustered trajectories. With the atlas prior included, more robust results are obtained, less sensitive to incorrect user settings.

5.5 Adjustable-Weight Prior (AWP)

The method with FWP employs spatial information from the atlas probabilistically. However, there is no control over the strength of the atlas. In this section, we extend our inference problem by considering a probability distribution over the atlas and then inferring its unknown parameters. We believe that the flexibility that this method offers on controlling the influence of the anatomical atlas has important applications in clinical studies. One such application is in analyzing pathological cases where fiber tracts might deviate significantly from normal cases and thus from an anatomical atlas. Since cluster centers are able to evolve during the course of the EM algorithm, the method is still able to cluster the fiber trajectories reasonably well when a small weight is given to the atlas. On the contrary, the algorithm might fail if the atlas is imposed strongly. So, the motivation here is to be able to control the strength of an

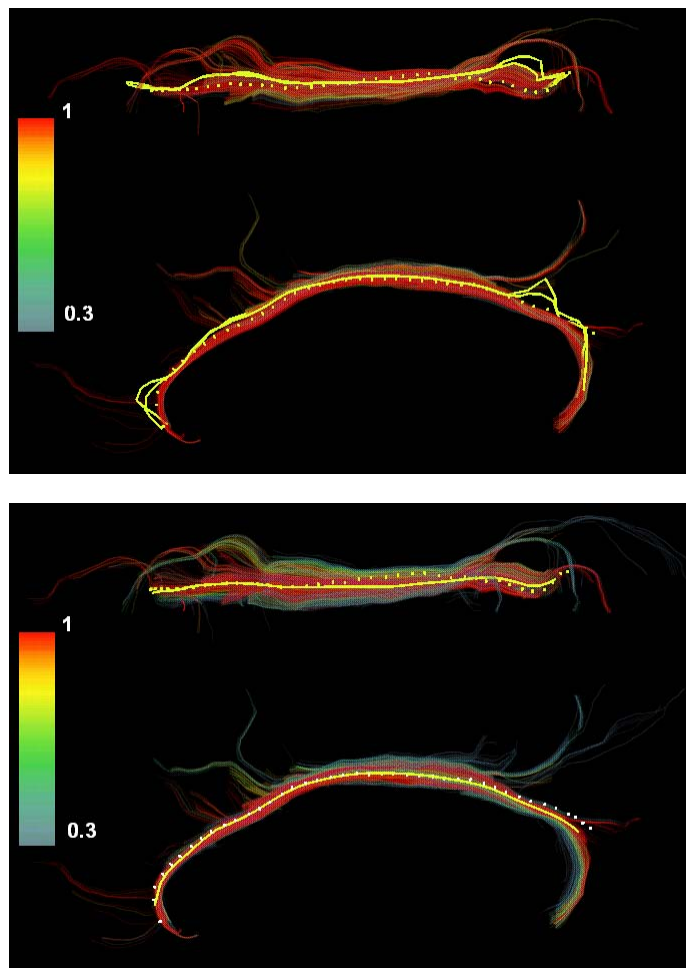


Figure 5-5: A comparison of the clustering results without (top) and with (bottom) an atlas. Axial and sagittal views are shown for superior cingulum. Trajectories are colored by their assigned membership probability. The cluster centers at consecutive EM iterations are shown in yellow, and dotted lines represent the initial centers. Without the atlas, and with improper setting of the clustering parameters the cluster centers drift and their extent increases as the algorithm proceeds. Less sensitivity to parameter setting and greater robustness is achieved with the atlas incorporated.

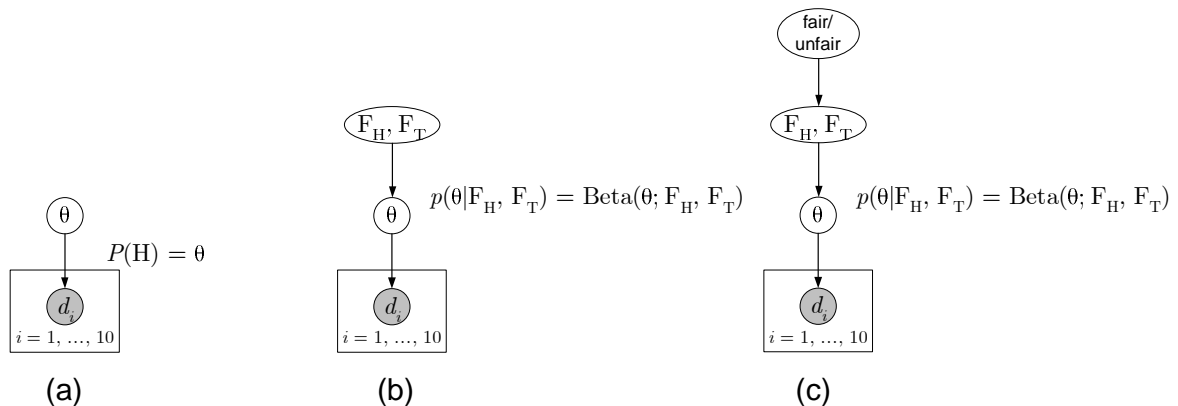


Figure 5-6: Graphical models for the coin flipping problem. (a) A coin is flipped 10 times and the probability of getting a Head, i.e., $P(H) = \theta$ is inferred from the outcome. (b) The model is extended to add our prior knowledge that most probably the coin is fair. The conjugate prior can be interpreted as some “fictitious data”, i.e., as if we had observed in the past that from $F_H + F_T$ tosses F_H came up Heads. (c) The model is further extended to add a model selection level; whether the coin is fair or not is also inferred from the coin tosses.

atlas by a single parameter. The influence of the anatomical prior in the proposed model reflects confidence in its accuracy and relevance. It can either be defined by the user or it can be inferred automatically, if multiple atlases are available. We first provide some background on hierarchical Bayes schemes and the general approach used to control the prior. Then, after a detailed description of our model for clustering, we demonstrate its properties through a set of experiments.

5.5.1 Hierarchical Bayes and Conjugate Prior

The hierarchical Bayes modeling approach has become very popular in recent years. It allows one to accommodate a richer class of models that can better capture the statistical understanding of a problem. The basic idea in a hierarchical model is that when we look at the likelihood function, and decide on the right priors, it may be appropriate to use priors that themselves depend on other parameters not incorporated in the likelihood. These parameters will require priors, which themselves may (or may not) depend on new parameters. Eventually the process terminates when we no longer introduce new parameters.

Let's consider a simple coin-flip example: We draw a coin from our pocket, and we want to estimate the probability of a Head after observing the outcome of ten tosses. Figure 5-6(a) depicts the graphical model for this simple experiment, where θ is the probability of Head in a single coin flip. If the observed data, D , is five heads and five tails, the estimate for θ is 50%, which is the solution of maximum likelihood estimation of the parameter θ :

$$\hat{\theta} = \arg \max_{\theta} p(D|\theta). \quad (5.7)$$

Now suppose that we tossed the coin ten times and it came out four Heads and six Tails. Our expectation is that the probability of a Head in the next toss is closer to 50% than 40%. After all, our intuition is that most probably it is a fair coin and thus $P(H) = 0.5$. Such prior knowledge can be integrated with the likelihood function, i.e., $p(D|\theta)$, using Bayes rule. For many standard distributions this can be done neatly using the concept of a conjugate prior.

A prior $p(\theta)$ is conjugate to a likelihood function $p(D|\theta)$ if the posterior has the same functional form as the prior. Different parameter values in the prior and posterior reflect the impact of observed data. Parameter values in the prior can be thought of as a summary of “fictitious observations” in some cases, which represent the strength of the prior (Figure 5-6(b)). To clarify how the strength of the prior is controlled, consider the coin tossing problem which has a binomial likelihood whose conjugate prior is a Beta distribution. Now, consider two priors $B(1000, 1000)$ and $B(5, 5)$, which both have an expected value of 0.5, while the former shows more confidence in the fact that the coin is fair. We can interpret this prior as if in the past we have tossed the coin 2000 times where Heads and Tails came up equally often. In the above coin-flipping example, where out of 10 tosses 4 came up Heads, the probability of the next toss to be a Head will be $1004/2010$, which is very close to 0.5. One can obtain the same result analytically by computing the posterior probability, $P(\theta|D)$, and seeing that it is a new Beta distribution with the parameters equal to the sum of the observed and the fictional counts of Heads.

Now, suppose you flip a coin 10 times and get all heads. Most probably the coin is not fair, but the model presented in Figure 5-6(b) cannot accommodate for proper model selection based on the outcome of the flipping trial. Figure 5-6(c) shows an extension of the graphical model where whether the coin is fair or not can be inferred from the coin flipping experiment. Such an extension provides a balance between the stability and flexibility of the model.

By this simple example, we aimed to provide an illustrative background on the strength of hierarchical Bayes modeling. We provide a detailed mathematical derivation in the context of our real clustering problem in the next section, in which we use a similar concept to control the effect of the atlas by using the Dirichlet distribution as the conjugate prior for the multinomial distribution.

5.5.2 Basic Definitions for AWP Modeling

Figure 5-7 shows the directed graphical model of the problem setup. Unlike the coin flipping example, we have a single trial experiment for each trajectory. We denote the unknown label of each trajectory by z_i , which is assumed to follow a single-trial multinomial distribution,

$$z_i \sim \text{Multinomial}(\boldsymbol{\pi}_i), \quad (5.8)$$

$$\Pr(z_i = k | \boldsymbol{\pi}_i) = \pi_{ik}, \quad (5.9)$$

and $\boldsymbol{\pi}_i$, the parameters of the multinomial distribution on z_i , follow a Dirichlet distribution, with parameters controlled by the atlas. Specifically, $\sum_{k=1}^K \pi_{ik} = 1$ and

$$\boldsymbol{\pi}_i \sim \text{Dirichlet}(\mathbf{q}_i) = \frac{\Gamma(q_{i0})}{\prod_{j=1}^K \Gamma(q_{ij})} \prod_{k=1}^K \pi_{ik}^{(q_{ik}-1)}, \quad (5.10)$$

where $\Gamma(\cdot)$ is the gamma function, and $q_{i0} = \sum_{k=1}^K q_{ik}$ and q_i is the parameter defined by the atlas. To control the influence of the atlas to some extent, \mathbf{q}_i can be set as $\mathbf{q}_i = \gamma \mathbf{q}'_i$ where $\gamma > 0$ is a weight factor and $\mathbf{q}'_i = [q'_{i1}, \dots, q'_{iK}]$ is the membership probability each trajectory takes from the atlas. Each q'_{ik} element denotes the atlas-specified membership of each trajectory i to cluster k . The membership probabilities

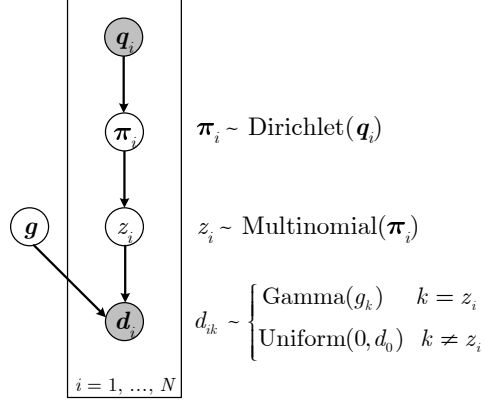


Figure 5-7: A graphical model that shows the inference of the priors and cluster parameters from the data when an atlas prior is available that describes the dependencies between different variables. $\boldsymbol{\pi}$: atlas prior, z : trajectory label, \mathbf{d} : data and \mathbf{g} : parameters of gamma distributions.

of each trajectory are then normalized, so that $\forall i, \sum_{k=1}^K q'_{ik} = 1$. The goal is to estimate the membership likelihood of each trajectory to each cluster based on the values of the d_{ik} 's and q_{ik} 's.

With such a setup, the expectation of the Dirichlet distribution for each trajectory is determined by the value it gets from the atlas and is independent of γ :

$$E[\pi] = \frac{q_i}{q_{i0}} = \frac{\gamma q'_i}{\gamma q'_{i0}} = \frac{q'_i}{q'_{i0}} \quad (5.11)$$

but the variance of the distribution is controlled by γ :

$$\text{Var}[\pi] = \frac{q_i(q_{i0} - q_i)}{q_{i0}^2(q_{i0} + 1)} = \frac{q'_i(q'_{i0} - q'_i)}{q_{i0}^2(\gamma q'_{i0} + 1)} \quad (5.12)$$

However, once the expectations, i.e., q'_{ik} 's, are set, the variation of the Dirichlet distribution with γ is limited as shown in Figure 5-8.

Since $\text{Dirichlet}([1 \ 1 \ \dots \ 1])$ is the uniform distribution, which translates to the minimum confidence in the provided information, better control is achieved with an alternative parameterization if a weighted average of the vector \mathbf{q}'_i and the unity vector is used, i.e.,

$$\mathbf{q}_i = a\gamma\mathbf{q}'_i + \mathbf{1}. \quad (5.13)$$

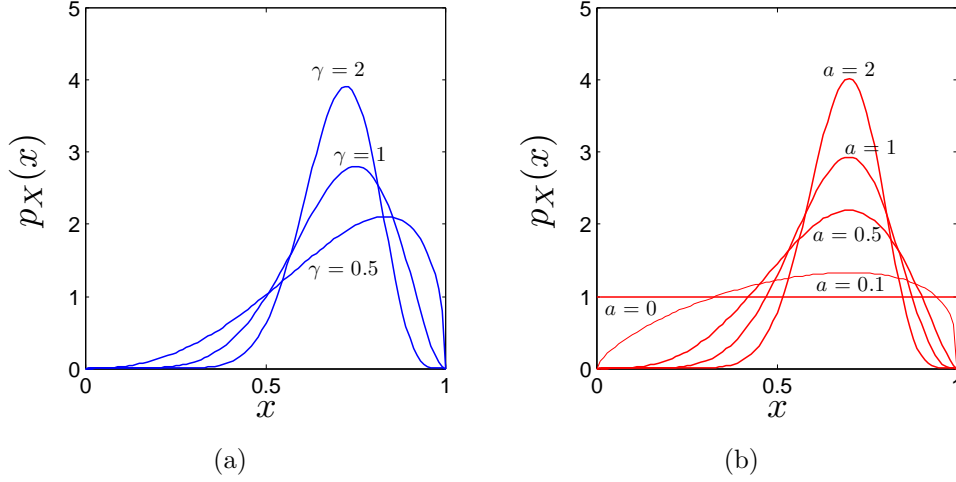


Figure 5-8: Control of the Dirichlet distribution with a weight factor. (a) With only a multiplying factor, γ , the expectation of the distribution is kept constant and only its standard deviation is varied. So, only a limited set of distributions can be used. (b) With a weighted averaging with the unity vector as in (5.13) the distribution takes any desired form from a uniform distribution ($a = 0$) to a highly peaked one ($a \gg 1$).

The weight, a , controls the degree of influence of the atlas on the clustering and can be set by the user. In the limiting case when $a = 0$, the Dirichlet distribution becomes a uniform distribution and hence the atlas does not have any control on the clustering. When $a = 1$ the model reduces to $\boldsymbol{\pi}_i \sim \text{Dirichlet}(\gamma \mathbf{q}'_i + \mathbf{1})$, which is slightly different from what we introduced earlier to guarantee that the parameters of the distribution are larger than unity and avoid a U-shaped distribution.

Alternatively, the weight a can be inferred from the atlas data if multiple atlases are available. One possible application of such a setting is to set the weight proportional to the correlation between the atlases, i.e., set a close to 1 when all atlases agree with each other and close to 0 when they disagree. Specifically, suppose that M atlases, $Q^{(1)}, \dots, Q^{(M)}$, are available, where $Q^{(m)} = \{\mathbf{q}_i^{(m)}\}_{i=1}^M$. A plausible setting for a can be obtained by forming the correlation matrix of $Q^{(m)}$'s and taking the average of its non-diagonal elements. The prior Q' is calculated as:

$$\mathbf{q}'_i = \frac{\sum_{m=1}^M \mathbf{q}_i^{(m)}}{M} \quad (5.14)$$

A more sophisticated possibility could combine multiple atlases, for example, if we

had a collection of voxelwise labellings (perhaps generated by hand), we could use STAPLE [94] to generate voxelwise label probabilities as input as described at the beginning of this section. When STAPLE combines labellings, it weights them according to how consistent they are with the others.

As detailed in Chapter 4, we assume that distances for each cluster follow a Gamma distribution with shape, α_k , and inverse scale parameters, β_k and we assume that each element of the vector \mathbf{d}_i follows the distribution

$$d_{ik} \sim \begin{cases} \text{Gamma}(g_k) & k = z_i \\ \text{Uniform}([0, d_0]) & k \neq z_i \end{cases}, \quad (5.15)$$

where g_k is the parameter set of the Gamma distribution, $\{\alpha_k, \beta_k\}$, for each cluster and $U(x; 0, d_0)$ is the uniform distribution function over $[0, d_0]$ with d_0 a large enough constant.

5.5.3 EM Formulation for AWP

As stated before, an EM approach is frequently used for inference with missing data or parameters. Again Z represents the hidden data, where $z_i = k$ represents the membership of the data point i in the cluster k . We denote the complete data likelihood by $p(D, Z|\Theta)$ where $\Theta = \{\mathbf{g}, \boldsymbol{\pi}\}$, i.e., the collection of the parameters of the gamma distribution functions, \mathbf{g} as well as parameters of the multinomial distribution, $\boldsymbol{\pi}$. The goal of our EM approach is to iteratively find the MAP estimates of the parameter Θ by maximizing the expectation of the log posterior, or equivalently the joint of the complete data at each iteration:

$$\Theta^{t+1} = \arg \max_{\Theta} \mathbb{E}_{Z|D, \Theta^t} [\log p(Z, D, \Theta)],$$

where superscript t denotes the set of parameters at iteration t . The maximization step takes the form:

$$\begin{aligned}
\Theta^{t+1} &= \arg \max_{\mathbf{g}, \boldsymbol{\pi}} \mathbb{E}_{Z|D, \mathbf{g}^t, \boldsymbol{\pi}^t} [\log p(Z, D, \mathbf{g}, \boldsymbol{\pi}; \mathbf{q})] \\
&= \arg \max_{\mathbf{g}, \boldsymbol{\pi}} \mathbb{E}_{Z|D, \mathbf{g}^t, \boldsymbol{\pi}^t} [\log p(D|Z, \mathbf{g}) + \\
&\quad \log p(Z|\boldsymbol{\pi}; \mathbf{q}) + \log p(\boldsymbol{\pi}; \mathbf{q}) + \log p(\mathbf{g})].
\end{aligned} \tag{5.16}$$

We assume a uniform distribution over \mathbf{g} , the parameters of the Gamma distribution, therefore the last term in the above equation is a constant and can be omitted from the maximization expression. Then:

$$\begin{aligned}
\Theta^{t+1} &= \arg \max_{\mathbf{g}, \boldsymbol{\pi}} \mathbb{E}_{Z|D, \mathbf{g}^t, \boldsymbol{\pi}^t} \left[\sum_{i=1}^N \log p(\mathbf{d}_i | z_i, \mathbf{g}) + \right. \\
&\quad \left. \sum_{i=1}^N \log p(z_i | \boldsymbol{\pi}; \mathbf{q}) + \log p(\boldsymbol{\pi}; \mathbf{q}) \right] \\
&= \arg \max_{\mathbf{g}, \boldsymbol{\pi}} \left[\log p(\boldsymbol{\pi}; \mathbf{q}) + \right. \\
&\quad \left. \sum_{i=1}^N \sum_{k=1}^K p_{ik} (\log p(\mathbf{d}_i | z_i = k, \mathbf{g}) + \log \pi_{ik}) \right].
\end{aligned} \tag{5.17}$$

With the assumption of independence of π_i 's, we write $\log p(\boldsymbol{\pi}; \mathbf{q}) = \sum_{k=1}^K \log p(\pi_i; q_i)$ where the π_i 's follow a Dirichlet distribution as specified in Eq. (5.10).

Expectation Step

The p_{ik} 's are computed in the expectation step using Bayes rule as follows:

$$\begin{aligned}
p_{ik} &= p(\mathbf{z}_i = k | \mathbf{d}_i, \mathbf{g}_k^t, \boldsymbol{\pi}_i^t) \\
&= \frac{p(\mathbf{d}_i | \mathbf{z}_i = k, \mathbf{g}_k^t) p(\mathbf{z}_i = k | \boldsymbol{\pi}_i^t)}{\sum_{j=1}^K p(\mathbf{d}_i | \mathbf{z}_i = j, \mathbf{g}_j^t) p(\mathbf{z}_i = j | \boldsymbol{\pi}_i^t)} \\
&= \frac{p(\mathbf{d}_i | \mathbf{z}_i = k, \mathbf{g}_k^t) \pi_{ik}^t}{\sum_{j=1}^K p(\mathbf{d}_i | \mathbf{z}_i = j, \mathbf{g}_j^t) \pi_{ij}^t}.
\end{aligned} \tag{5.18}$$

Maximization Step

Now we only look at the i 'th component of Θ :

$$\begin{aligned}\Theta_i^{t+1} = \arg \max_{\mathbf{g}, \boldsymbol{\pi}_i} & \left(\sum_{k=1}^K (q_{ik} - 1) \log \pi_{ik} + \right. \\ & \left. \sum_{k=1}^K (p_{ik} \log p(\mathbf{d}_i | z_i = k, \mathbf{g}) + p_{ik} \log \pi_{ik}) \right),\end{aligned}\tag{5.19}$$

where Θ_i is the collection of parameters of the multinomial distribution $\boldsymbol{\pi}_i$ and parameters of the Gamma distribution \mathbf{g} . We do the maximization with respect to each parameter separately as follows:

Updating parameters of the multinomial distribution

In order to get the updated parameters of the multinomial distribution:

$$\begin{aligned}\boldsymbol{\pi}_i^{(t+1)} &= \arg \max_{\boldsymbol{\pi}_i} \sum_{k=1}^K \left((q_{ik} - 1) \log \pi_{ik} + p_{ik} \log \pi_{ik} \right) \\ &= \arg \max_{\boldsymbol{\pi}_i} \sum_{k=1}^K \left((q_{ik} + p_{ik} - 1) \log \pi_{ik} \right),\end{aligned}\tag{5.20}$$

under the constraint that

$$\sum_{k=1}^K \pi_{ik} = 1 \quad \forall i\tag{5.21}$$

To do the maximization, we add the Lagrange multiplier, λ , to Equ.(5.20) and differentiate with respect to π_{ik}

$$\begin{aligned}\boldsymbol{\pi}_i^{(t+1)} &= \arg \max_{\boldsymbol{\pi}_i} \sum_{k=1}^K \left((q_{ik} + p_{ik} - 1) \log \pi_{ik} \right) \\ &\quad - \lambda \left(\sum_{k=1}^K \pi_{ik} - 1 \right)\end{aligned}\tag{5.22}$$

or

$$\frac{q_{ik} - 1 + p_{ik}}{\pi_{ik}} - \lambda = 0 \quad (5.23)$$

This results in:

$$\pi_{ik} = \frac{q_{ik} - 1 + p_{ik}}{\lambda}. \quad (5.24)$$

λ is obtained by summing up the Equ.(5.24) over the clusters:

$$\lambda = \frac{\sum_{k=1}^K (q_{ik} - 1 + p_{ik})}{\sum_{k=1}^K \pi_{ik}} = q_{i0} - K + 1. \quad (5.25)$$

Substituting the values from Equ. (5.13) for q_{ik} in the above equation results in the following expression which explicitly shows the relation to the parameters controlling atlas influence:

$$\pi_{ik} = \frac{a\gamma q'_{ik} + p_{ik}}{a\gamma + 1}. \quad (5.26)$$

In the limiting case when $a = 0$, the above equation reduces to $\pi_{ik} = p_{ik}$, and when $a\gamma \gg 1$ it becomes $\pi_{ik} = q'_{ik}$.

Updating parameters of the gamma distribution

The updated cluster parameter \mathbf{g} is obtained from:

$$\mathbf{g}^{t+1} = \arg \max_{\mathbf{g}} \sum_{i=1}^N \sum_{k=1}^K p_{ik} \log p(\mathbf{d}_i | z_i = k, \mathbf{g}) \quad (5.27)$$

The maximization equations are computed by differentiating the following expression from the EM formulation with respect to each parameter:

$$\begin{aligned} Q(\mathbf{g}) &= \sum_{i=1}^N \sum_{k=1}^K p_{ik} \left(\log \text{Gamma}(d_{ik}; \mathbf{g}_k) \right. \\ &\quad \left. + \log \prod_{j \neq k} \text{U}(d_{ij}; 0, d_0) \right). \end{aligned} \quad (5.28)$$

The inverse scale parameter is obtained by $\partial Q(\mathbf{g})/\partial \beta_k = 0$, which results in:

$$\sum_{i=1}^N p_{ik} \left(\frac{\alpha_k}{\beta_k} - d_{ik} \right) = 0 \quad (5.29)$$

or

$$\beta_k = \alpha_k \frac{\sum_{i=1}^N p_{ik}}{\sum_{i=1}^N p_{ik} d_{ik}}. \quad (5.30)$$

The shape parameter is given by $\partial Q(\mathbf{g})/\partial \alpha_k = 0$, thus:

$$\begin{aligned} \sum_{i=1}^N p_{ik} \left[\log d_{ik} + \log \alpha_k + 1 + \log \frac{\sum_{i=1}^N p_{ik}}{\sum_{i=1}^N p_{ik} d_{ik}} + \right. \\ \left. d_{ik} \frac{\sum_{i=1}^N p_{ik}}{\sum_{i=1}^N p_{ik} d_{ik}} - \psi(\alpha_k) \right] = 0 \end{aligned} \quad (5.31)$$

where $\psi(.) = \Gamma'(.)/\Gamma(.)$ is the digamma function. The resulting equation

$$\log \alpha_k - \psi(\alpha_k) = -\log \frac{\sum_{i=1}^N p_{ik}}{\sum_{i=1}^N p_{ik} d_{ik}} - \frac{\sum_{i=1}^N p_{ik} d_{ik}}{\sum_{i=1}^N p_{ik}}, \quad (5.32)$$

does not have a closed-form solution. However, as previously, good approximations can be obtained by noting that

$$\log \alpha_k - \psi(\alpha_k) \approx \frac{1}{\alpha_k} \left(\frac{1}{2} + \frac{1}{12\alpha_k + 2} \right). \quad (5.33)$$

This gives:

$$\alpha_k \approx \frac{3 - x + \sqrt{(x - 3)^2 + 24x}}{12x}, \quad (5.34)$$

where

$$x = \log \left(\frac{\sum_i p_{ik} d_{ik}}{\sum_i p_{ik}} \right) - \frac{\sum_i p_{ik} \log(d_{ik})}{\sum_i p_{ik}}. \quad (5.35)$$

Similar to FWP, we update the cluster centers and recompute the distance vectors, once the EM algorithm converges.

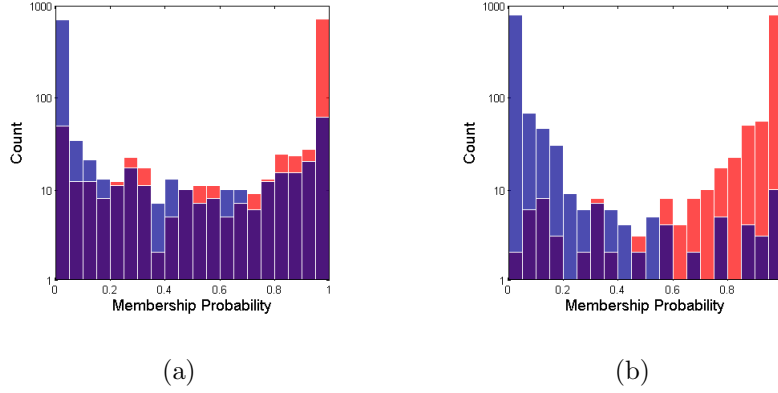


Figure 5-9: A comparison of the clustering results with two degrees of atlas influence and with an atlas prior that agrees with the membership likelihood inferred from the mixture model: (a) $a = 0$, i.e., no atlas, and (b) $a = 1$. Histograms of the membership probability p_{ik} , are shown for synthetic data colored by the true cluster assignment. The misclassification is indicated by the presence of red samples on the left side of the histograms or blue samples on the right. The ratio of mis-clustered data points decreases from 15% to less than 2% when the atlas is used. Note that the y-axis is in logarithmic scale.

5.5.4 Results

To show that the proposed method is able to effectively control the influence of the prior information on the clustering, we apply it to simulated data. For illustration purposes, we consider two clusters and construct \mathbf{d} by drawing samples from gamma and uniform distributions as specified in (5.15). Two scenarios were examined: In the first one, the atlas priors, q'_{ik} s, were drawn from a uniform distribution in the $[0.8, 1]$ range for each data point i that belongs to the k th cluster. This models a case where the atlas prior is in agreement with the membership probability inferred from the data only. In the second case, q'_{ik} s were drawn from a uniform distribution in the $[0, 0.2]$ range if the data point i belongs to the k th cluster, i.e. an extreme case where the atlas priors oppose the membership probability inferred from the data.

Figure 5-9 shows a histogram of the posterior probabilities, colored based on the true clustering and for the case where the atlas prior is in agreement with the membership likelihood inferred from the data. Clustering results for two different values of $a = 0$, i.e. no atlas, and $a = 1$, i.e. full atlas control are shown. The ideal

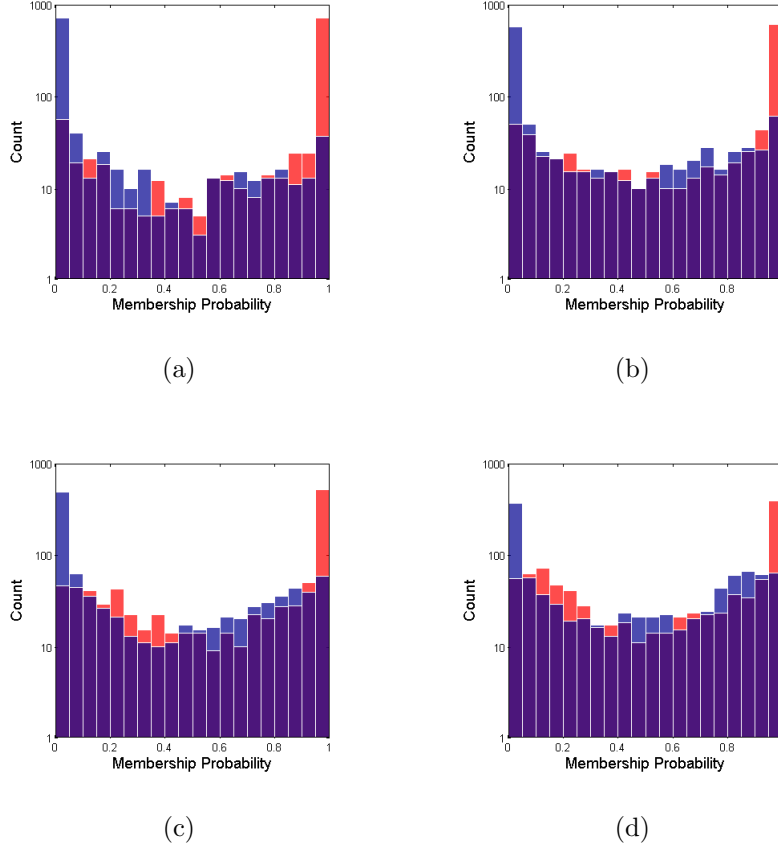


Figure 5-10: A comparison of the clustering results with different degrees of atlas influence, but with an atlas prior that disagrees with the membership likelihood inferred from the mixture model: (a) $a = 0$, i.e., no atlas, (b) $a = 0.25$, (c) $a = 0.5$, and (d) $a = 1$. Histograms of the membership probability, p_{ik} s, are shown for synthetic data colored by the true cluster assignment. Misclassification is indicated by the presence of red samples on the left hand side of the histograms. The misclustering ratio increases as the influence of the atlas is increased. Note that the y-axis is in logarithmic scale.

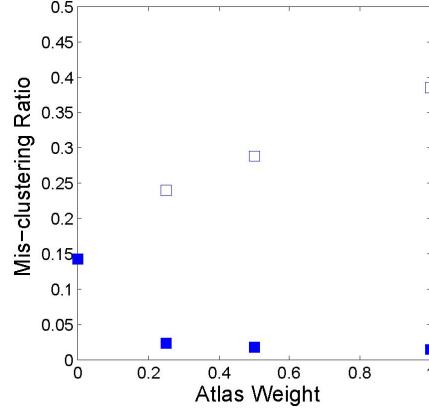


Figure 5-11: Effect of atlas weight, a , on the mis-clustering ratio and when the atlas prior agrees (filled symbols) or disagrees (open symbols) with the membership likelihood inferred from the mixture model.

clustering is defined as a case where all data points are assigned to their true clusters, in other words when all data points from each cluster appear in either the upper or lower half of the histogram. While without the atlas roughly 15% of the data points are mis-clustered, only a few data points are mis-clustered once the atlas is used.

Figure 5-10 shows the case where the atlas disagrees with the membership likelihoods inferred from the data and for different values of the weight a . In this case the number of mis-clustered data points increases as the atlas weight is increased. Figure 5-11 summarizes the impact of the atlas weight on the number of the mis-clustered data points. When the atlas agrees with the data, the mis-clustering ratio decreases as the atlas weight is increased and when it disagrees with the atlas the mis-clustering ratio increases.

To demonstrate the effectiveness of our method in successfully clustering the fiber trajectories in such a case, Figure 5-12 shows the clustering results for trajectories from the cingulum. When the atlas is imposed strongly some of the trajectories that belong to cingulum, i.e., are similar in shape and are located close to other trajectories, are rejected only because they do not have enough overlap with the atlas ROI. With a proper choice of the atlas weight these trajectories are also included in the results.

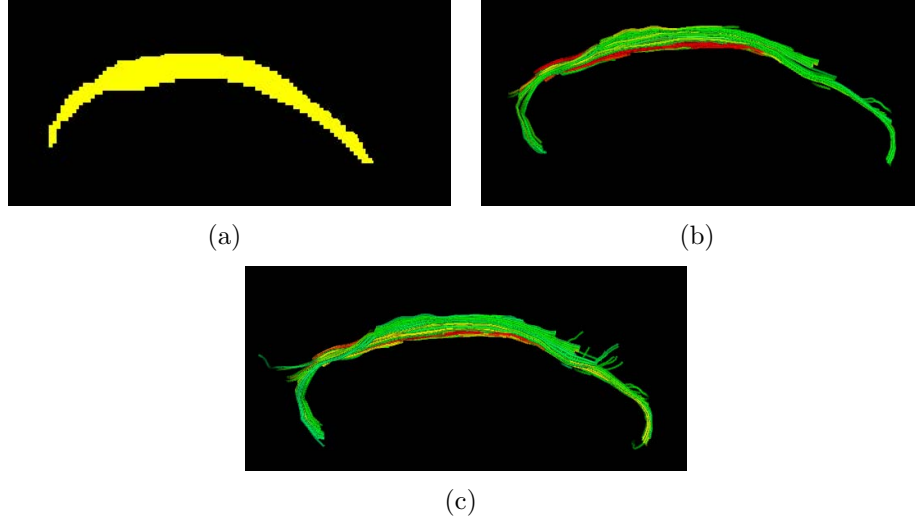


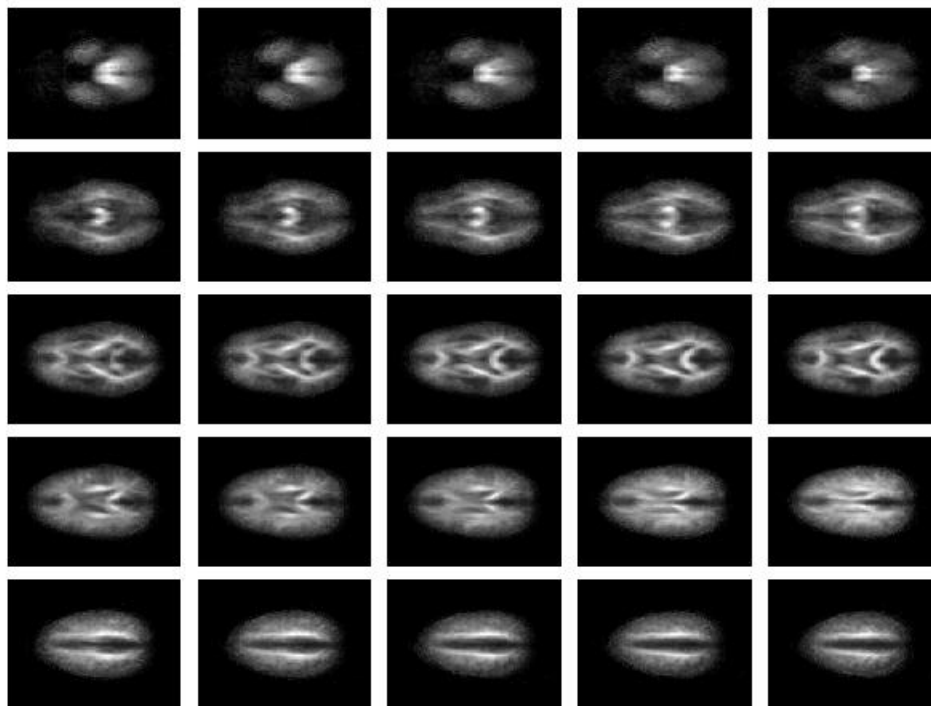
Figure 5-12: The influence of atlas weight on the clustering of cingulum trajectories: (a) atlas-specified ROI for cingulum and clustering results with (b) $a = 1$ and (c) $a = 0.5$. When the atlas is imposed strongly some of the trajectories that belong to the cingulum are rejected only because they do not have enough overlap with the atlas ROI.

5.6 Registration

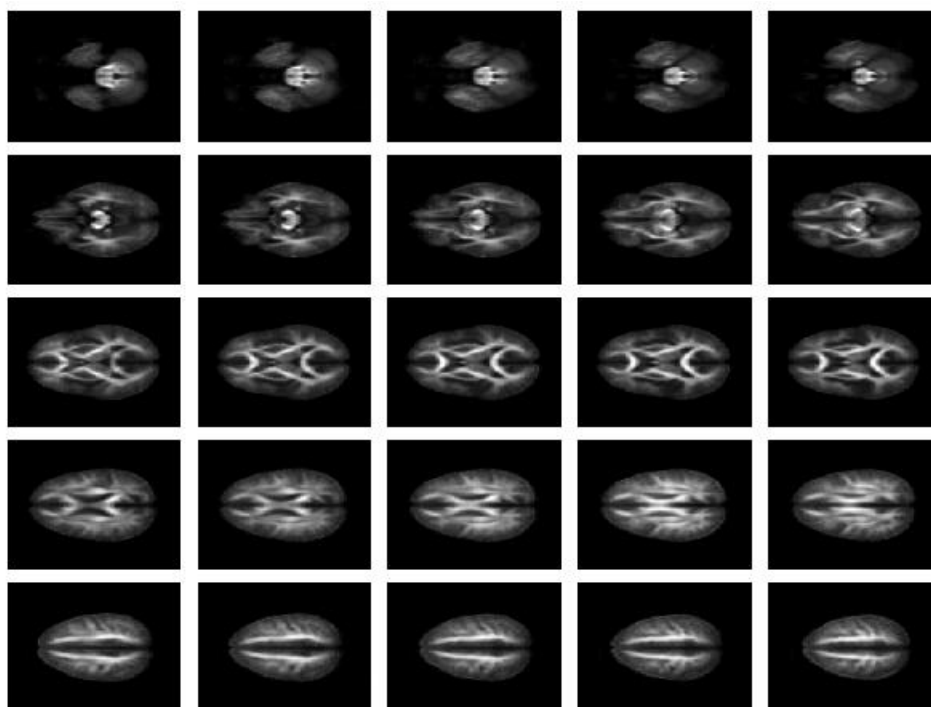
In population studies and in particular in the approaches presented in this chapter that use an anatomical atlas, spatial alignment of the data is an important step. In this section we present the details on the steps of our registration approach and investigate the sensitivity of clustering to errors in alignment of the data.

In the population studies carried out in this thesis, we first apply the stochastic congealing method [104] on the corresponding FA volumes of the subjects, which simultaneously aligns group of images to a template derived from the subjects so that every member of the population approaches the central tendency of the collection at the same time. This approach avoids introducing bias in creating the template image. Figure 5-13 shows the mean volume before and after congealing of 44 subjects with a 9-parameter transformation (translation, rotation, and scale).

The template FA volume produced by the congealing algorithm is then aligned to the FA volume of the MNI atlas by running a 9-parameter affine registration [95]. Figure 5-14 illustrates the composite images of the FA volumes of the atlas with the congealing output before and after running the pair-wise registration. Misalignments



(a)



(b)

Figure 5-13: Demonstration of the performance of the congealing algorithm used to align the FA volumes from 44 subjects to a common space. Axial slices of the mean FA volumes are shown (a) before and (b) after registration. Note the improved sharpness of the plots in (b).

are shown in red in both plots. As can be seen in the figure, the registration works quite well on the data.

Figure 5-15 shows the impact of registration error on the clustering results. A comparison is made between our clustering algorithm with FWP and an algorithm which assigns membership probabilities only based on the atlas information. The mean square error (MSE) in membership probabilities is plotted as a function of the translation and rotation errors in part (a) and (b) of the figure, respectively. For both cases the clustering error grows almost quadratically with the registration error. However, the clustering error for our algorithm is almost half of that from the atlas-only method. This figure shows that our method is more robust to error in registration compared to an approach where only the atlas is used.

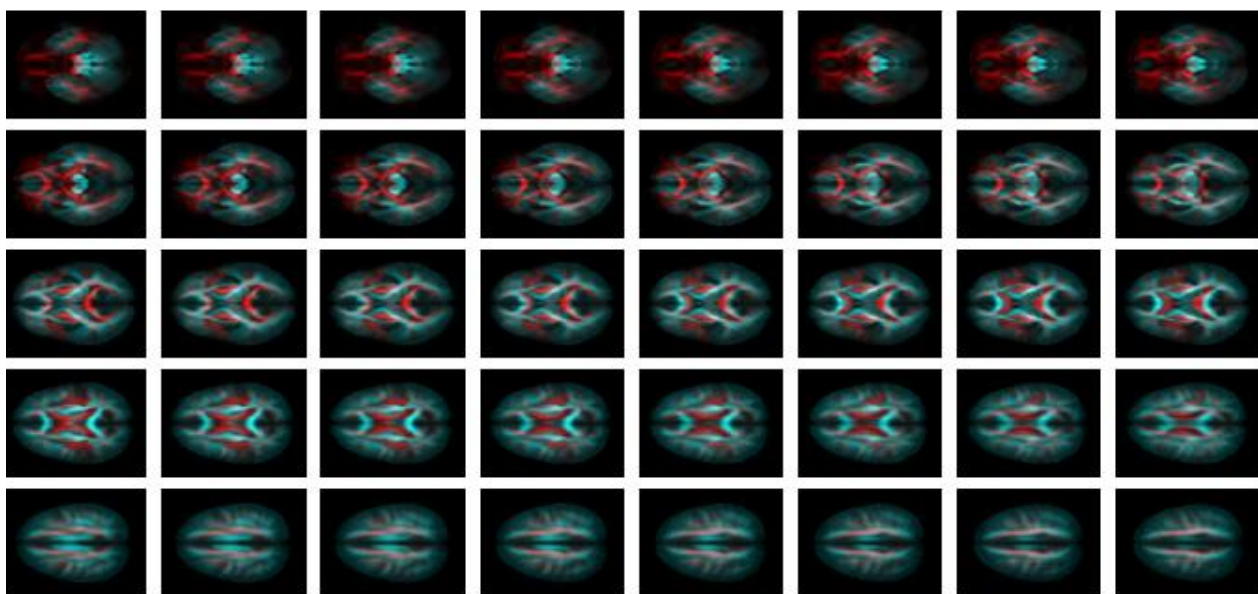
Note that quantitative analysis of FA or any other feature along the trajectories is only affected by the errors in computation of posterior membership probabilities, as the FA values or other features of interest at each voxel of trajectories are stored during the tractography phase, not after registration.

5.7 Summary

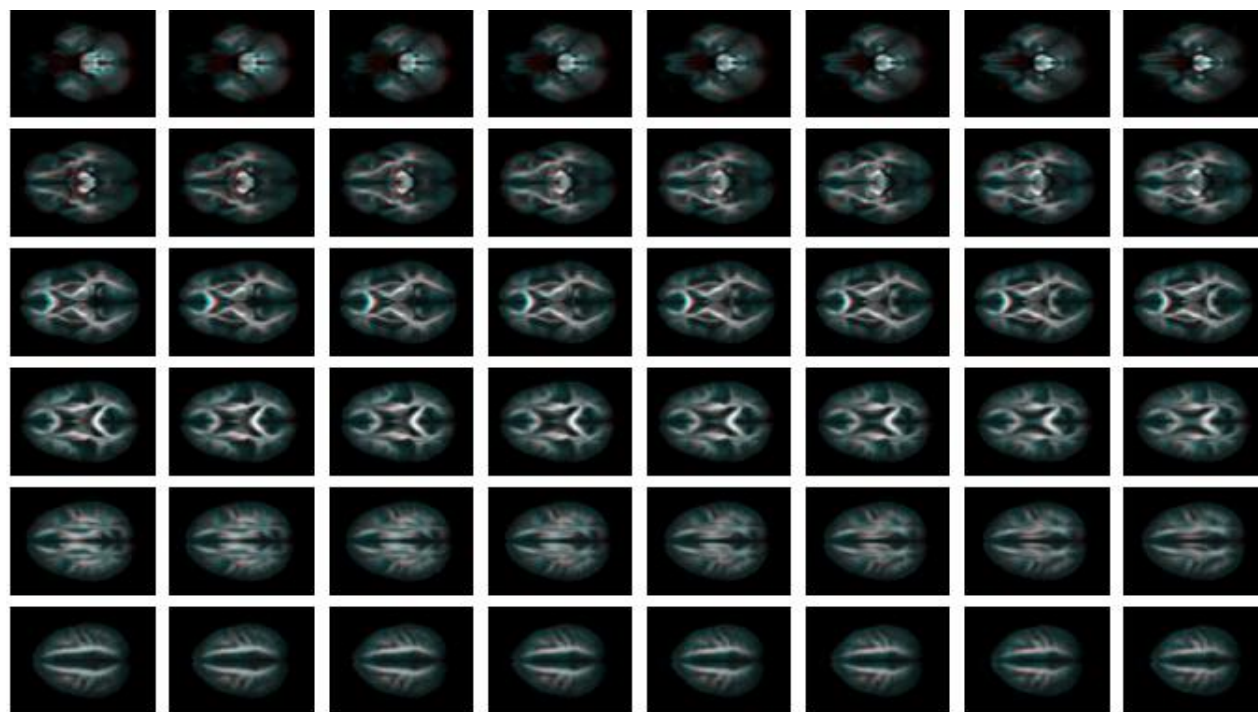
In this chapter, two novel methods were introduced to incorporate anatomical knowledge in the clustering of white matter fiber trajectories.

In the first method, called FWP, we extended our expectation-maximization (EM) clustering algorithm, presented in Chapter 4, to cluster the trajectories in which an atlas served as the fixed prior on the labels. The atlas guides the clustering algorithm and makes the resulting bundles anatomically meaningful. In addition, the atlas provides the seed points for the tractography as well as the initial settings of the EM algorithm.

In the second method, called AWP, we extended the FWP approach to be able to weight the prior. The Dirichlet distribution as a conjugate to the multinomial distribution was used to model the prior information so that influence of the atlas can be controlled by a parameter. This parameter can be either set by the user



(a)



(b)

Figure 5-14: Demonstration of the performance of the affine registration used to map the output of the congeling algorithm to the MNI space. After overlapping the images, misalignment is shown in red (a) before and (b) after applying the affine registration.

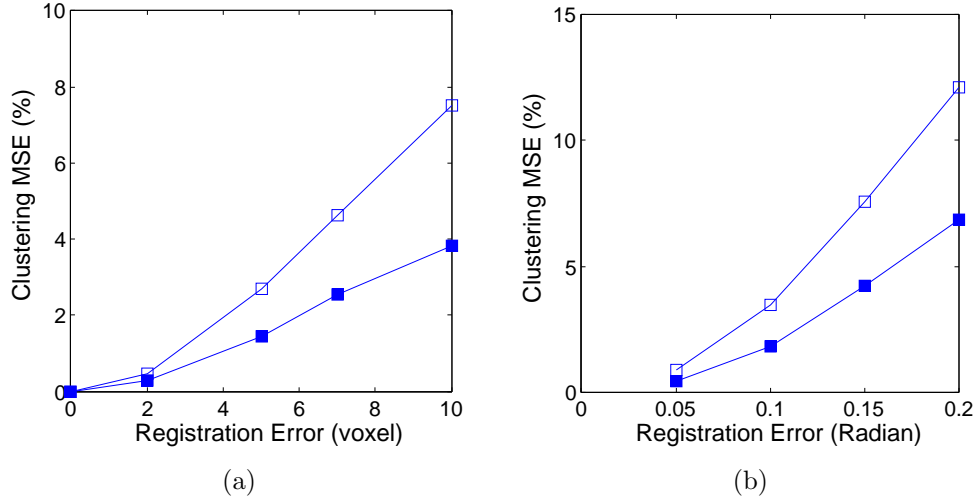


Figure 5-15: Sensitivity of the clustering results to registration errors. Mean square error of the posterior probabilities are plotted as a function in the (a) translation and (b) rotation error. Filled symbols are results of our algorithm, whereas open symbols correspond to an atlas-only clustering method.

or inferred from the atlas information if more than one atlas is present to reflect the degree of confidence in the prior knowledge. Results presented in this chapter demonstrate the effectiveness of the proposed approach. To our knowledge this is the first implementation which offers direct control over the strength of the anatomical information for the clustering of fiber trajectories.

Chapter 6

Experimental Results on Clinical Data

This chapter presents some of the clinical applications of the methods developed in the previous chapters. The first section presents the application of our model for efficient 3-D visualization of fiber tracts in neurosurgery. The next section reports preliminary results that demonstrate the capabilities of our clustering and quantitative analysis method to study brain development. A significant portion of this chapter is dedicated to the quantitative analysis that we performed on fiber bundles that are hypothesized to be related to schizophrenia. These bundles include the splenium and genu parts of the corpus callosum, uncinate fasciculus, inferior longitudinal fasciculus, inferior occipito-frontal fasciculus, and cingulum. Some of the findings reported in this chapter confirm previously published reports, whereas the rest of the findings are new and revealed by our tract-oriented quantitative analysis.

6.1 Three-Dimensional Modeling and Visualization of Bundles

Real-time visualization of fiber tracts is beneficial for neurosurgical applications to minimize post-operative neurological deficits while maximizing tumor removal [87].

Trajectories of the fiber tracts, which are extracted from DT-MRI data using a tractography method, are being visualized along with anatomical data from structural MRI for this purpose. Although this provides a viable representation of the underlying connectivity between functional regions, the complexity of the structure limits the usefulness of the method for surgical application. From a surgical point of view, the spatial extent of specific fiber tracts would suffice in order to avoid damage during resection. A hull that surrounds all fibers of a certain tract has been proposed to represent the extent of the tract for this purpose [31, 33]. However, information enclosed by individual trajectories is lost when constructing the hull. Such information is valuable when registering pre-operative data onto intra-operative acquisitions or when performing quantitative analysis during the post-operative course. In order to preserve an aggregate of the information for each fiber tract, it is necessary to obtain the point correspondence between individual trajectories that belong to that tract. As discussed in Chapter 3, finding the point correspondence of 3-D curves in a large dataset is not trivial and that is why this task has not been rigorously tackled in the past.

As a byproduct of the proposed clustering method presented in Chapter 4, a spatial model of the fiber bundles represented by the mean trajectory and its spatial variation is also obtained. This is shown in Figure 6-1 in which the abstract models of five fiber bundles are visualized by their spatial mean and iso-surfaces corresponding to the mean plus three standard deviations (3σ) of the 3-D coordinates calculated along the cluster center by the method presented in Section 4.5.2. Such an abstract spatial model for fiber bundles could be used for neurosurgery applications.

Our models enable one to easily visualize the extent of the fiber tracts adjacent to the brain lesions to minimize the damage to the bundles when removing the lesion. Conventional DT-MRI-based approaches usually visualize a map of the fractional anisotropy [50] or bundle segmentations based on the FA map. However, such approaches are not accurate at fiber crossings where FA is low. An alternative approach is to visualize the fiber trajectories [44], but rendering such a large dataset may be difficult in real-time applications.

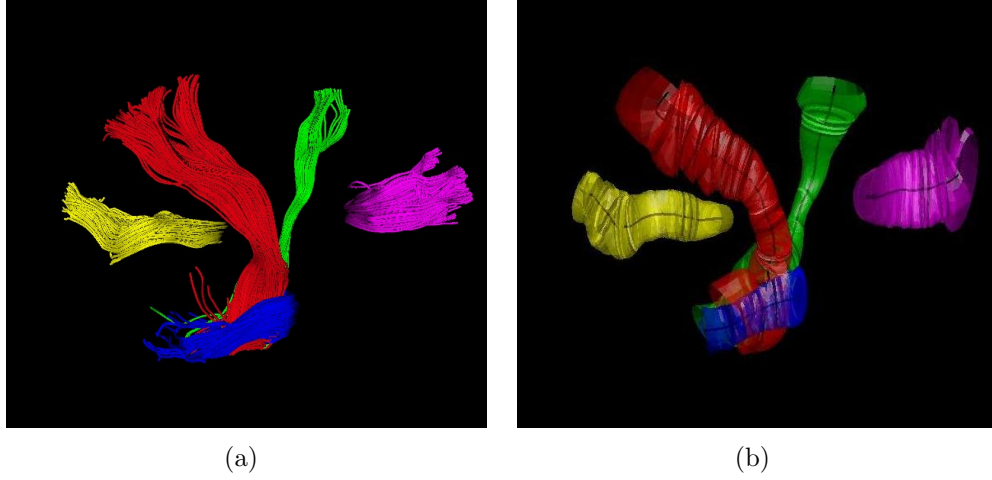


Figure 6-1: (a) Trajectories of 5 different clusters used for quantitative analysis: splenium (yellow), corticospinal (red), corticobulbar (green), middle cerebellar peduncle (blue), and genu (magenta). (b) A model representation of the bundles as the mean trajectory and the isosurfaces corresponding to spatial variation of the clusters.

Having the mean trajectory for each fiber tract allows us to study the shape of the bundle which can be used to study either brain development or the changes that occur in shape of the bundles because of diseases. As an example, the curvature versus the normalized arc length of the cluster centers are plotted in Figure 6-2 for each cluster shown in Figure 6-1. For a 3-D curve, \mathbf{r} , parametrized as a function of the arc length, the curvature and torsion are defined as:

$$\kappa = \frac{\|\mathbf{r}' \times \mathbf{r}''\|}{\|\mathbf{r}'\|^{3/2}} \quad (6.1)$$

and

$$\tau = \frac{(\mathbf{r}' \times \mathbf{r}'') \cdot \mathbf{r}'''}{\|\mathbf{r}' \times \mathbf{r}''\|^2}, \quad (6.2)$$

respectively, where $'$ denotes differentiation with respect to the arc length. The B-spline representation of the fiber trajectories allows us to evaluate the derivatives analytically and avoid further numerical errors.

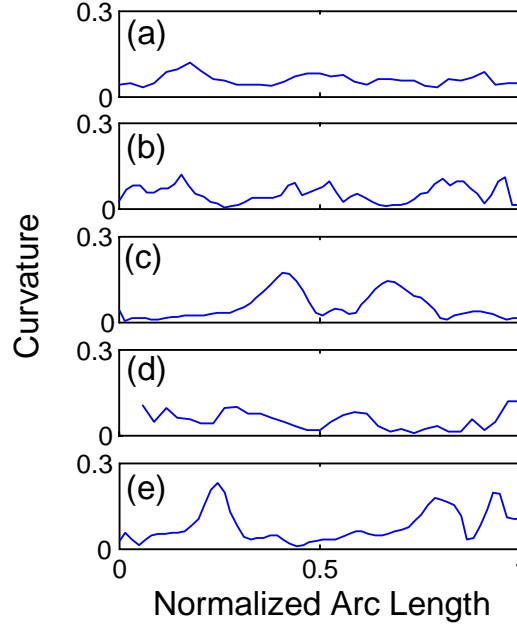


Figure 6-2: Curvature of the cluster center along its normalized arc length for fiber bundles shown in Figure 6-1: (a) splenium, (b) genu, (c) middle cerebellar peduncle, (d) corticospinal, and (e) corticobulbar fiber tracts.

6.2 Investigation of Neonate Brain Development

Investigation of early brain development is of great scientific and clinical importance. Prematurely born infants are vulnerable to brain injuries and early detection of abnormalities may help their treatment and care. Diffusion tensor MRI enables visualization and quantification of white matter pathways well before other imaging techniques. However, studies performed so far have not fully benefited from the potentials of this modality. Prior work on neonate DT-MRI analysis has been limited to the comparison of scalar diffusion parameters, such as fractional anisotropy (FA), in manually defined regions of interest (ROIs) [13,42]. As discussed in Chapter 2, such approaches are sensitive to the accuracy and reproducibility of the ROIs specified by the experts. Errors in defining the ROIs are especially critical in analyzing neonate DTI due to the limited resolution and SNR. More importantly, spatial patterns of the tract maturation are lost in ROI-based quantitative analysis. Tract-oriented analysis, wherein the fiber tract acts as a common coordinate system, enables unambiguous

quantification of the spatial information. To our knowledge, this is the first tract-oriented analysis of the neonate DTI data where mapping of the fiber pathways to the common coordinate system is performed rigorously. Dubois *et al.* [32] proposed a similar approach, wherein the quantitative parameters are either reported at manually selected landmarks along the fiber pathways or averaged over the whole fiber bundle.

6.2.1 Material and Method

Very low birth weight infants born before 30 weeks postmenstrual age underwent structural and diffusion MRI at approximately 32 weeks and approximately 42 weeks postmenstrual age. Imaging was carried out without sedation and with informed consent utilizing a protocol approved by the institutional review board. Whole brain DTI required 10 minutes of scanning time on 1.5 T GE scanner with 31 directions, 6 baseline images, and (0.7, 0.7, 1.5) mm resolution. Manual outlining of ROIs for tractography was performed in 3D Slicer(www.slicer.org). The Expectation Maximization algorithm method, presented in Chapter 4, was used to cluster the trajectories in our mixture model framework and to obtain the point correspondence between the trajectories in each cluster. The common coordinate system is obtained as proposed in Chapter 3 by constructing a distance map from each cluster center. The parameters of interest are averaged over all trajectories that belong to a cluster and plotted along the common coordinate system.

6.2.2 Results

Clustered white matter trajectories for cortico-spinal, cingulum, and uncinate fasciculus are shown in Figure 6-3 (a)-(c) and (d)-(f) at two time points of 32 and 42 weeks postmenstrual age for one of the subjects. FA-colored trajectories clearly show the spatial patterns of the tract development, especially in the cortico-spinal tract. Box-plots of the FA variation along the normalized tract arc length of the superior part of the cingulum and at the above two time points are shown in Figure 6-4. As can

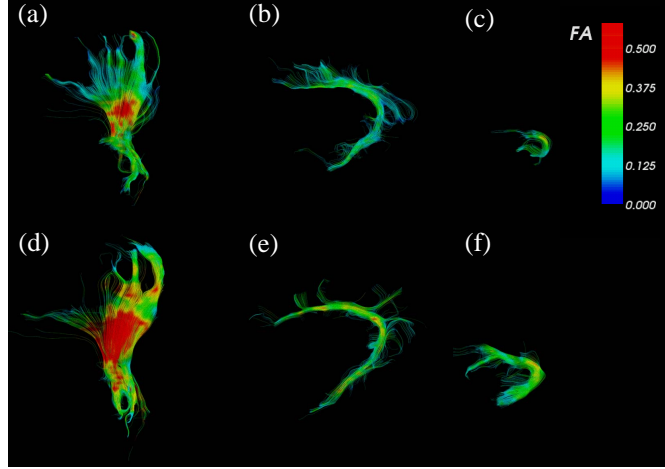


Figure 6-3: FA-colored trajectories from (a), (d) cortico-spinal, (b), (e) cingulum and (c), (f) uncinate fasciculus at 32-wk (up) and 42-wk (down) postmenstrual age. Spatial patterns of the tract development are clearly seen.

be clearly seen, only the posterior part of the tract exhibits a significant increase in the fractional anisotropy. Whereas tract-oriented analysis shows 76% increase in the mean FA (from 0.17 to 0.30) in the posterior part of the superior cingulum, ROI-based analysis on the entire superior cingulum shows only 18% FA increase (from 0.22 to 0.26). This observation signifies the importance of analyzing the spatial dependence of the FA changes enabled by our tract-oriented approach.

6.3 DT-MRI Findings in Schizophrenia

Schizophrenia is one of the major disabling brain disorders. It has been hypothesized that oligodendroglial dysfunction and subsequent myelin abnormalities contribute to the schizophrenic syndrome [27]. Diffusion tensor MRI has been employed by several groups in the past to investigate myelin integrity in patients with schizophrenia. Reduced diffusion anisotropy has been reported in prefrontal cortex [18], cingulum [86], uncinate fasciculus [19, 49], corpus callosum [1, 4, 36], and arcuate fasciculus [19]. However, other studies find no significant difference in some of the above structures [36, 37, 84, 86]. We believe that the inconsistency in experimental results is in part due to the difference in specifying the regions over which the anisotropy is measured.

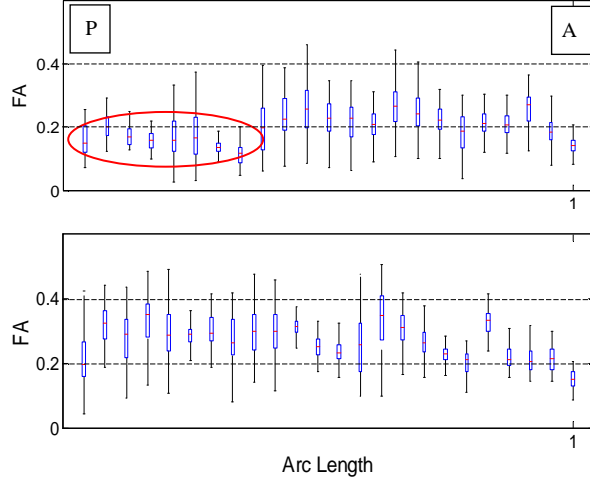


Figure 6-4: Box-plot of the FA variation along the tract arc length for the superior part of the cingulum and at 32-wk (up) and 42-wk (down) postmenstrual age. Only the posterior part shows a significant FA increase. ROI-based analysis fails to detect such spatial dependencies.

Clinical studies in the past are mostly based on either averaging the quantitative parameters over expert-specified regions of interest (ROIs) or using a voxel-by-voxel comparison. As mentioned in Chapter 1, the former is sensitive to the accuracy of specifying the ROI, while the latter is susceptible to registration errors.

Most recently, a tract-specific analysis has been proposed to more accurately define the measurement regions [46]. Jones *et al.* perform tractography to obtain the pathways of each fiber tract and then report a single number by averaging the parameter of interest along all pathways that belong to a given tract. However, local variation of the quantitative parameters is not captured in their analysis.

In this section, we present the result of our tract-oriented analysis in which parameters of interest are studied along the arc length of the tract. This allows us to study local variations that are missed in an ROI-based analysis or in the study performed in [46]. We apply the proposed procedure on a population of 18 patients with chronic schizophrenia and 19 normal subjects, analyzing several fiber bundles that have been indicated in schizophrenia research, and demonstrating the sensitivity of our approach to identifying the group differences in diffusivity measures such as fractional anisotropy (FA) and mean diffusivity (MD).

6.3.1 Method

In this study, the probabilistic tract-oriented quantitative analysis described in Chapter 5 is used. We use an atlas of fiber tracts [92], which is composed of a set of labeled ROIs, each corresponding to an anatomically-known bundle of fiber tracts in the human brain. These ROIs are used to seed the tractography and to guide the clustering algorithm. As mentioned earlier, we add prototype curves to the atlas to represent the shape of the trajectories in each tract. This is a one-time process performed by an expert, by manual selection of trajectories that are representative of the shape of the tracts. These prototype curves are also used as the initial cluster centers in the clustering algorithm. Fractional anisotropy maps calculated from DT-MRI data are first mapped into a common coordinate system using a congealing registration [103] and then into the atlas (MNI) space using an affine registration [95].

Atlas-specified regions corresponding to each fiber tract are mapped into the coordinate system of each case using the mapping parameters obtained from registration. These mapped regions are then dilated to ensure that they contain the fiber tract of interest. These regions are used to seed a streamline tractography [10], as described in Algorithm 2.1 (in Chapter 2). The quantitative parameters of interest, such as fractional anisotropy and the diffusivity eigenvalues, are collected at each point on the trajectories and stored for further quantitative analysis.

Trajectories obtained from the tractography algorithm in each case are then mapped into the common MNI space, and then clustered using our EM algorithm with the fixed-weight prior, presented in Section 5.4. We employ the atlas-specified ROIs as the spatial prior information, while the prototype trajectories of the atlas are used as the initial cluster centers for initializing the EM iteration. As described in Section 5.4, the membership likelihood from the mixture model is combined with the atlas prior in a Bayesian framework to determine the membership probability. The output of the clustering algorithm is the probabilistic assignment of the trajectories to each cluster, a set of cluster centers, and the point correspondence between the trajectories and the cluster centers.

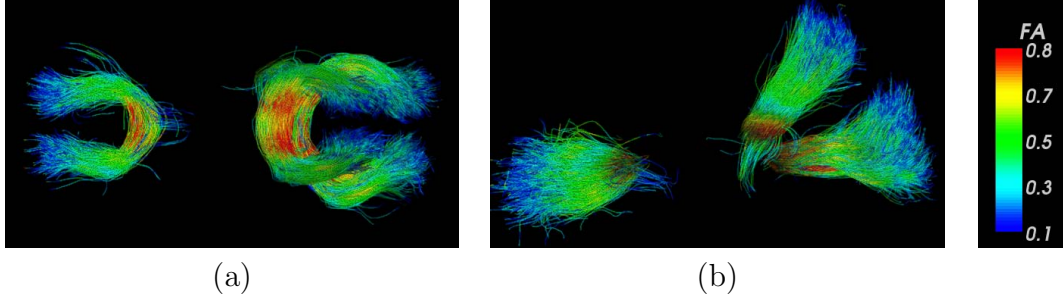


Figure 6-5: (a) Axial and (b) sagittal view of the clustered trajectories from genu and splenium parts of the corpus callosum. Trajectories from all cases are registered into the atlas space, clustered, and colored with the local fractional anisotropy.

Tract-oriented quantitative analysis is performed by calculating the weighted averages of the parameters of interest along the arc length of the cluster centers. By doing so, trajectories with low membership probability contribute less to the quantitative analysis. Such analysis can be performed on a case-by-case basis or on all cases that belong to a group. Analysis of variance (ANOVA) and permutation testing (with 1000 permutations) are used for statistical group analysis.

6.3.2 Experiments and Results

Here, we present the results of our tract-oriented quantitative analysis performed on a population of schizophrenia and control subjects. We examine corpus callosum fiber tracts, left and right uncinate fasciculus, inferior longitudinal fasciculus, inferior occipito-fronto fasciculus, and cingulum bundles.

For these experiments, we used DTI images acquired on a 3T GE system using an echo planar imaging (EPI) DTI sequence. We used a double echo option to reduce eddy-current related distortions, and an 8-Channel coil that allows us to perform parallel imaging using ASSET (Array Spatial Sensitivity Encoding Techniques, GE) with a SENSE-factor (speed-up) of 2 to reduce the impact of EPI spatial distortions. Eighty-five axial slices parallel to the AC-PC line covering the whole brain were acquired, with the following parameters: 51 directions with $b = 900$, 8 baseline scans with $b = 0$, $TR = 17000$ ms, $TE = 78$ ms, $FOV = 24$ cm, 144×144 encoding steps,

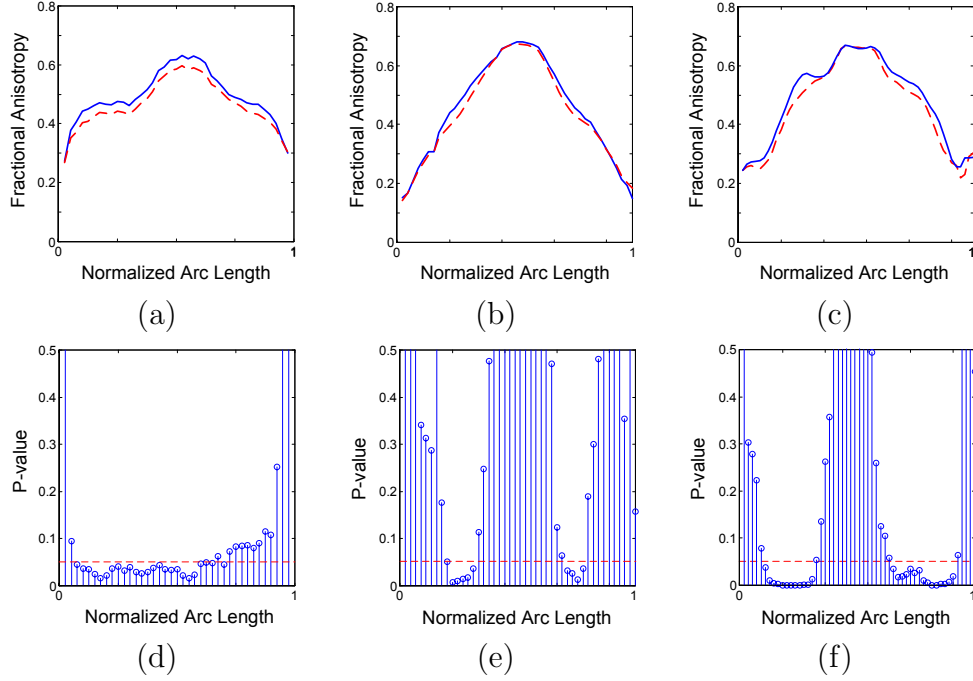


Figure 6-6: The variation of FA along the arc length, from left to right of the brain, for genu (a), upper splenium (b) and lower splenium (c) for healthy (solid lines) and diseased (dashed lines) subjects and the corresponding p-values. Significant differences are observed on the left side of the genu and the lateral portion of the splenium.

and 1.7 mm slice thickness.

Corpus Callosum

Some of the schizophrenia symptoms are hypothesized to be related to the disruptions of the inter-hemisphere connectivity, which makes study of the corpus callosum of great interest. To analyze the corpus callosum in schizophrenia, we analyzed DTI data from 19 control subjects and 18 chronic schizophrenics, matched on age, gender and parental socioeconomic status. Figure 6-5 shows the clustered trajectories from genu and splenium of the corpus callosum, colored based on the local fractional anisotropy. Our method allows us to divide the splenium into its upper and lower parts (tracts interconnecting parietal and occipital lobes respectively), as they have different shapes. Figure 6-6 compares the fractional anisotropy in genu and upper and lower splenium for healthy and schizophrenic subjects. Significant FA reduction

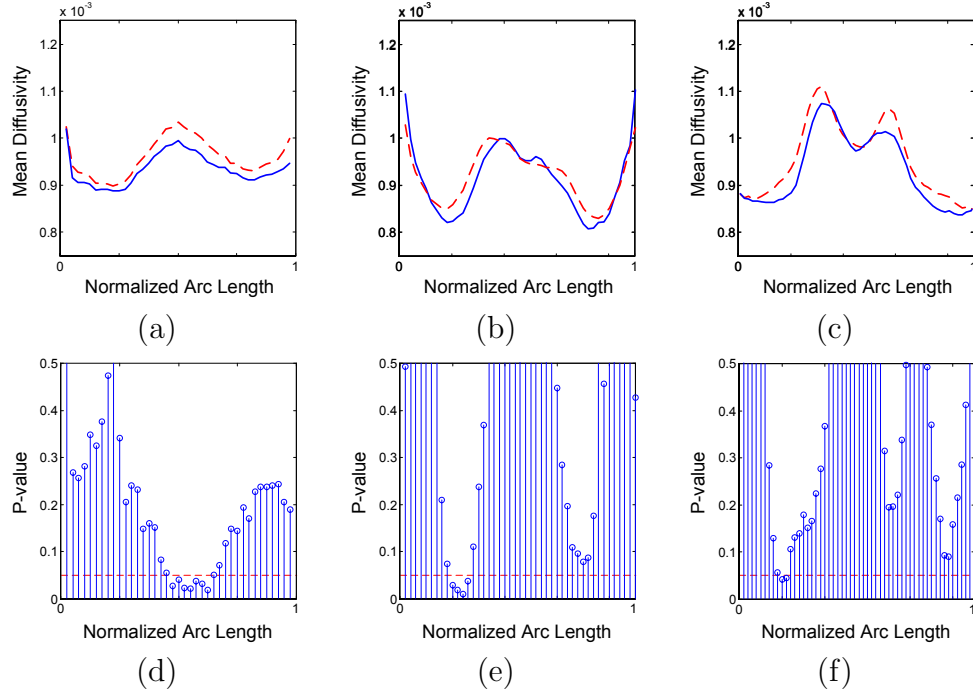


Figure 6-7: Similar to Figure 6-6 but for mean diffusivity along the tract.

is observed in genu, which confirms earlier reports of similar observations [68]. Moreover, our analysis clearly indicates that FA reduction is primarily observed in the left part of the genu, which is consistent with a few detailed voxel-based studies [68].

We also observed significant reduction of the fractional anisotropy in the splenium, especially in its lower part. As can be seen in Figure 6-6, the FA reduction is seen bilaterally, in the middle of each side. To our knowledge, only one voxelwise analysis [4] was sensitive enough to observe such a bilateral FA reduction in the splenium. Similar analysis is performed for mean diffusivity and the results are given in Figure 6-7. Increased mean diffusivity is significant only in small portions of the genu (in the middle) and the splenium (in the left).

Uncinate Fasciculus(UF), Inferior Longitudinal Fasciculus (ILF), and Inferior Fronto-Occipito Fasciculus (IFO)

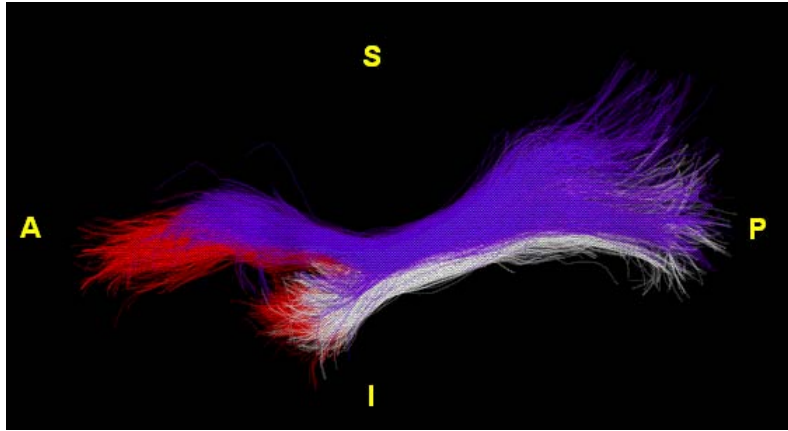
Besides corpus callosum, fronto-temporal connections are the second most frequently indicated fiber tracts in pathophysiology of schizophrenia. In this work, we thus

examine the uncinate fasciculus, which is the most prominent of all white matter fiber tracts connecting the frontal and temporal lobes, and which has been implicated in schizophrenia in several publications [19]. We also study the inferior longitudinal fasciculus and inferior fronto-occipito fasciculus bundles. From this experiment on, we used 44 subjects, 22 control subjects and 22 chronic schizophrenics, matched on age, gender and parental socioeconomic status. Figure 6-8 shows the trajectories of three bundles of UF, ILF and IFO in the atlas space and the successful separation of these three bundles by our clustering algorithm. Note that in the second plot we shifted the bundles spatially away from each other for a better visualization of the bundles. In part (c) of the figure, the shifted bundles are colored by the FA values.

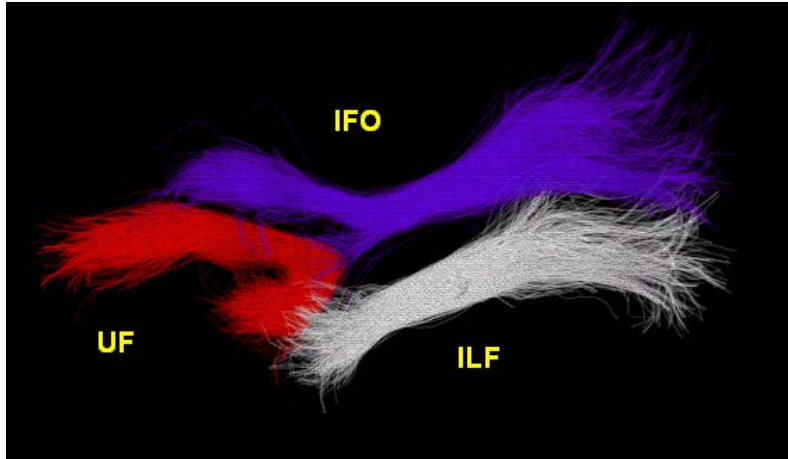
Regarding the Uncinate Fasciculus, as shown in Figure 6-9, an increase in FA is significant for a small part at the middle of both left and right UF bundles. We do not observe statistically significant differences in the mean diffusivity, the parallel or perpendicular diffusivity. However, as shown in Figure 6-10 we observe a trend of increase in parallel diffusivity for the left bundle and a trend of decrease in perpendicular diffusivity for the right bundle. This observation suggests that two different processes might be responsible for an increase in the FA value of the left and right bundles. This observation, which has not been reported in the past, indicates that FA alone is not enough for understanding what is changing in the brain. Although the origins of increased parallel diffusivity are unknown and it opposes general views of axonal damage, it should be noted that such an increase is not impossible. In fact, increased parallel diffusivity has been reported previously for displaced fiber tracts [80].

To study changes in brain asymmetry in schizophrenia, in Figure 6-11 we have plotted the fractional anisotropy against the normalized arc length of the uncinate fasciculus in control and schizophrenia subjects. In both groups, significant left-greater-than-right asymmetry is observed in the inferior part of the uncinate fasciculus, while the asymmetry is reversed in the superior part.

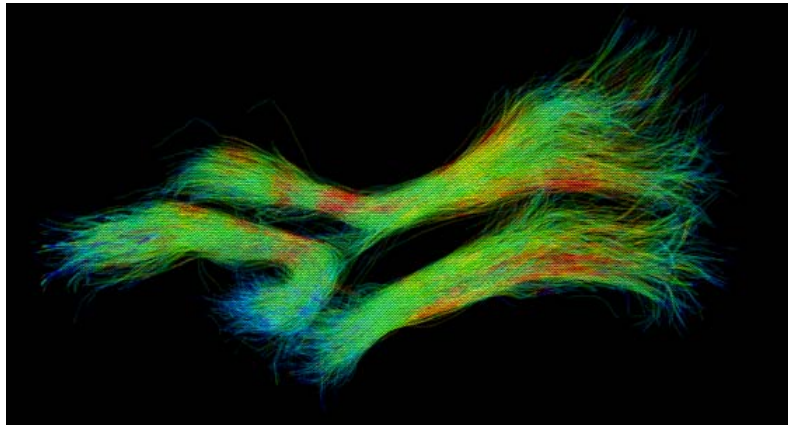
Similar analysis performed on the inferior longitudinal fasciculus (ILF), shown in Figure 6-12, reveals that no significant change in the fractional anisotropy is observed



(a)



(b)



(c)

Figure 6-8: (a)The trajectories of three bundles of UF, ILF and IFO in the atlas space are clustered by our algorithm. These bundles are manually shifted in space for better visualization in (b) and are colored by FA values in (c).

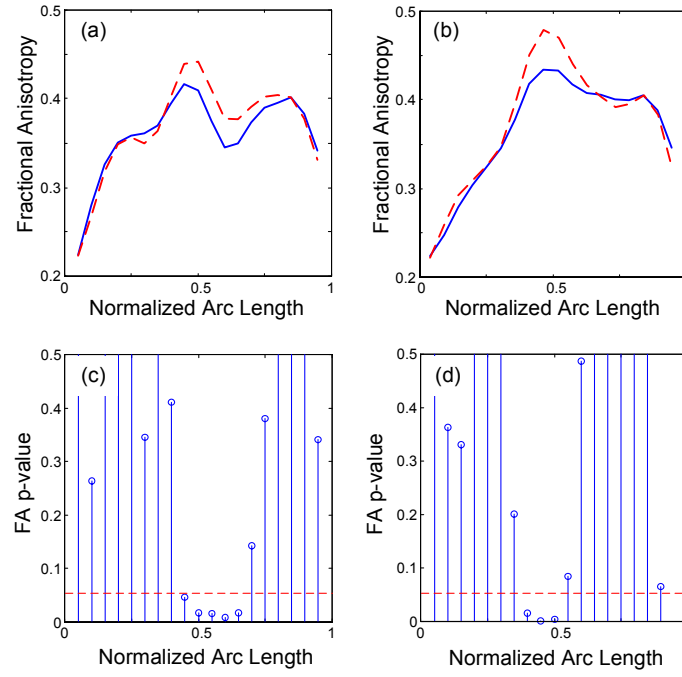


Figure 6-9: Variation of fractional anisotropy along the tract arc length from inferior to anterior for (a) left and (b) right uncinate fasciculus in healthy (solid lines) and diseased (dashed lines) subjects and the corresponding p-values (c) and (d).

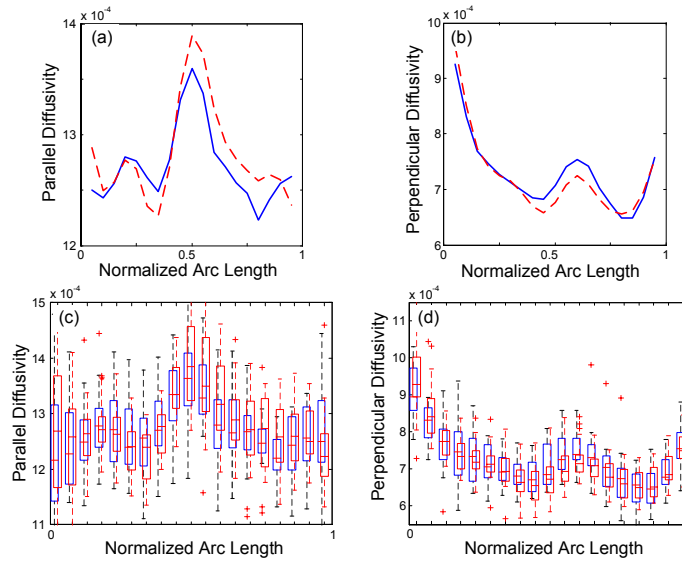


Figure 6-10: Variation of (a) parallel diffusivity and (b) perpendicular diffusivity along the tract arc length from inferior to anterior for (a) left and (b) right uncinate fasciculus in healthy (solid lines) and diseased (dashed lines) subjects and the corresponding box plots (c) and (d).

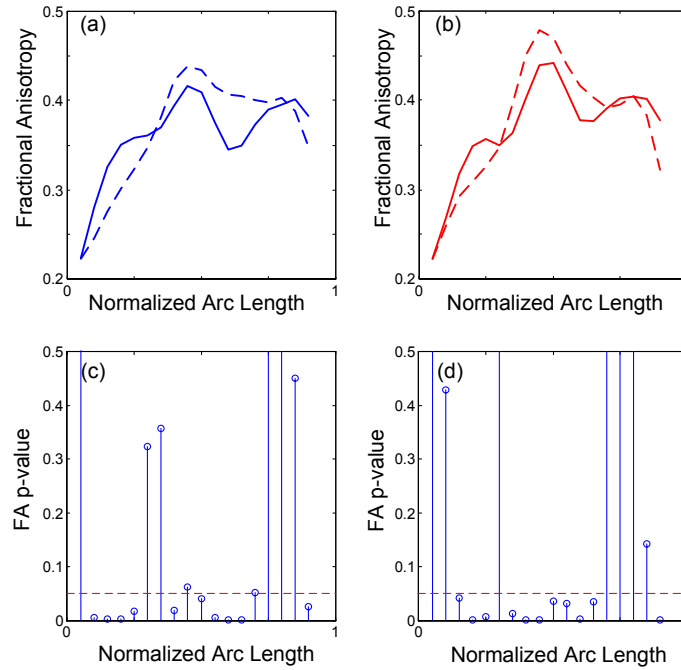


Figure 6-11: Comparison of the fractional anisotropy along the normalized arc length of the left (solid lines) and right (dashed lines) UF and for the (a) control and (b) schizophrenia cases. The corresponding p-values in (c) and (d) demonstrate statistically significant left-greater-than-right asymmetry in the inferior part and right-greater-than-left asymmetry in the superior part of the UF in both groups.

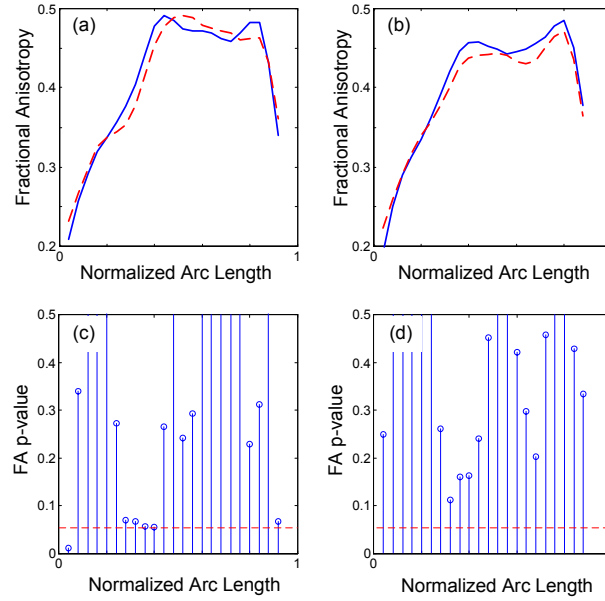


Figure 6-12: Comparison of the fractional anisotropy for (a) left and (b) right ILF for control (solid lines) and schizophrenia (dashed lines) groups, along with the corresponding p-values (c) and (d). No significant change in FA is observed for these bundles.

in this structure. Reduced FA is observed in part of the left ILF, which is mainly due to increased perpendicular diffusivity, as shown in Figure 6-13(a). There is some evidence of decreased parallel diffusivity in the right ILF, as seen in Figure 6-13(b), although these changes are not statistically significant. Further study of this structure, perhaps with a larger dataset, is needed to arrive at statistically meaningful conclusions.

Figure 6-14 shows the fractional anisotropy along the arc length of the left and right inferior occipito-fronto fasciculus (IFO) for the control and schizophrenia groups as well as the corresponding p-values. Reduced FA is observed in two small parts of the left IFO. Further analysis shows that this is mainly due to increased perpendicular diffusivity at these two sites. On the other hand, a combination of increased FA (in one site) and decreased FA (in two sites) of the right IFO is observed. Detailed analysis presented in Figure 6-15 demonstrates that significant increase in the perpendicular diffusivity is observed in the right IFO, but is not reflected fully in the FA analysis since the parallel diffusivity is also increased in some portions of this structure. Again,

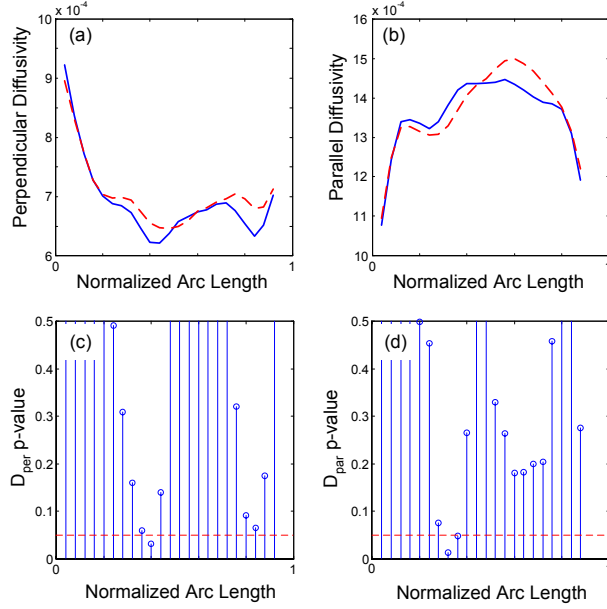


Figure 6-13: Comparison of the (a) perpendicular diffusivity in the left ILF and (b) parallel diffusivity in the right ILF for the healthy (solid lines) and schizophrenia (dashed lines) groups, along with the corresponding p-values (c), (d).

this analysis demonstrates that a quantitative analysis based only on the changes in FA does not completely capture changes in the white matter integrity.

In Figure 6-16 the samples of cluster centers for left and right UF, ILF and IFO bundles are visualized. The size of each sample is proportional to the significance of the difference between the FA value of the schizophrenics and controls at that point. We colored the increase and decrease in FA with red and blue respectively.

Cingulum

The cingulum is a C-shaped fiber tract that is part of the limbic system and handles functions related to memory and emotion. The cingulum bundle connects the cingulate gyrus to other regions in the limbic lobe, especially the hippocampus, which is involved in memory formation. Based on its function, the cingulum fiber tract is of great interest to schizophrenia studies.

Figure 6-17 illustrates the clustered left and right cingulum bundles in different views. These bundles are colored by their local FA values. Figure 6-18 shows the fractional anisotropy along the arc length of the superior part of the left and right

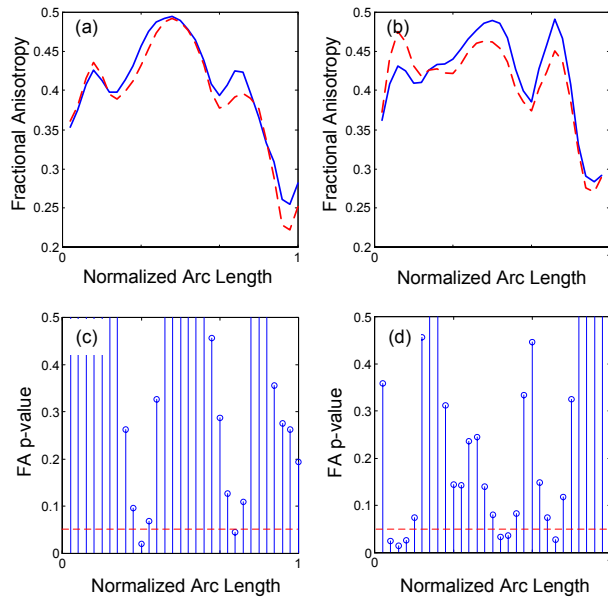


Figure 6-14: Comparison of the fractional anisotropy for the (a) left and (b) right IFO and for the healthy (solid lines) and schizophrenic (dashed lines) groups, plotted against the normalized arc length of the bundle. The corresponding p-values are shown in (c) and (d).

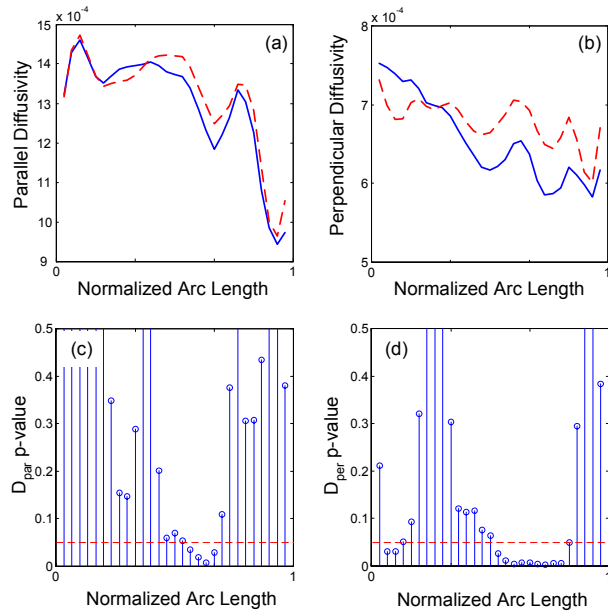
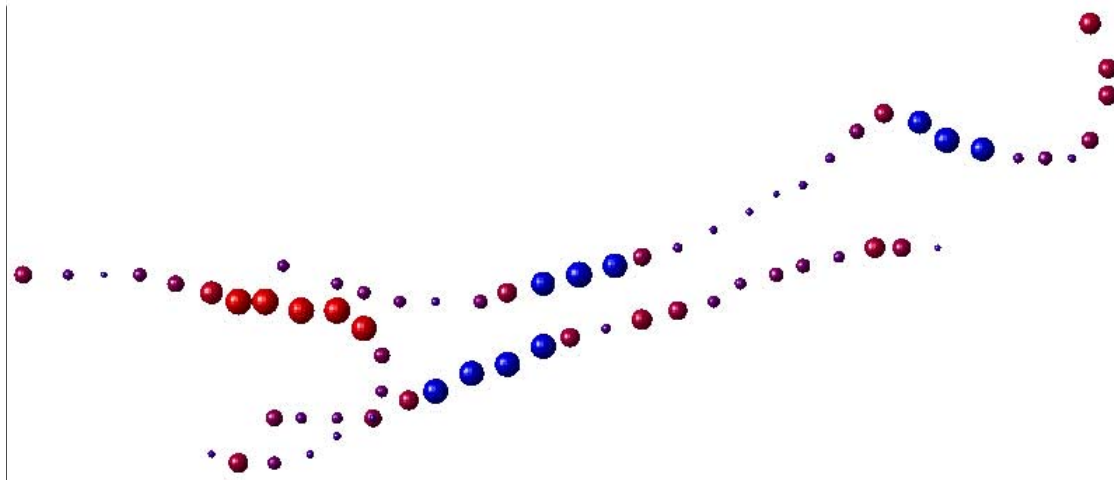
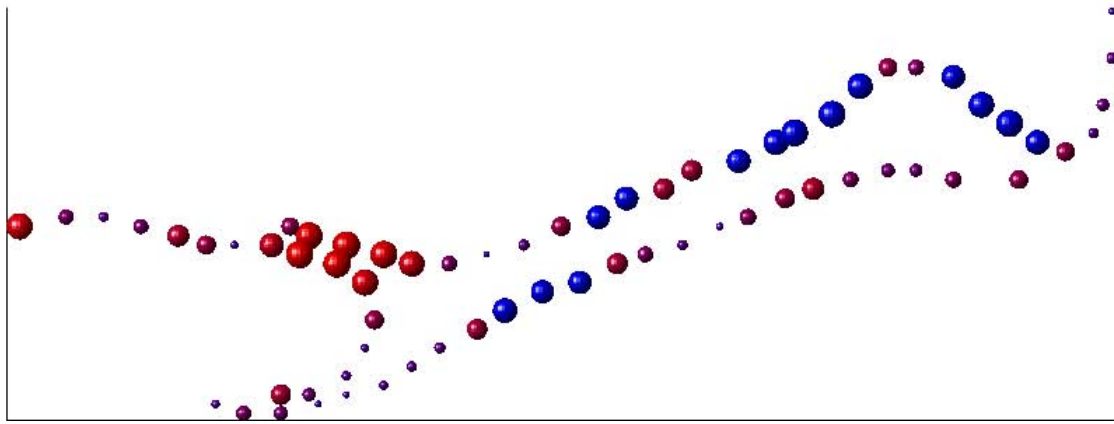


Figure 6-15: Comparison of the (a) parallel and (b) perpendicular diffusivity for the right IFO in healthy (solid lines) and schizophrenia (dashed lines) groups. Significant increase in the perpendicular diffusivity is observed for a large portion of this bundle.



(a)



(b)

Figure 6-16: Visualization of samples of cluster centers for (a) left and (b)right UF, ILF and IFO bundles. The radius of each sample is proportional to the significance of the difference between the FA value of the schizophrenics and controls at that point. The increase and decrease in FA is highlighted in red and blue respectively.

cingulum and the corresponding p-values. For the left cingulum, increased FA is observed in the middle of the superior cingulum. Further analysis demonstrates that this is due to a combination of increased parallel diffusivity and decreased perpendicular diffusivity in schizophrenia cases. On the other hand, the right cingulum exhibits reduced FA in its anterior part. As shown in Figure 6-19, both parallel and perpendicular diffusivity increase in the anterior part of the right cingulum. Hence, the FA plot in Figure 6-18 does not completely reflect changes in the right cingulum.

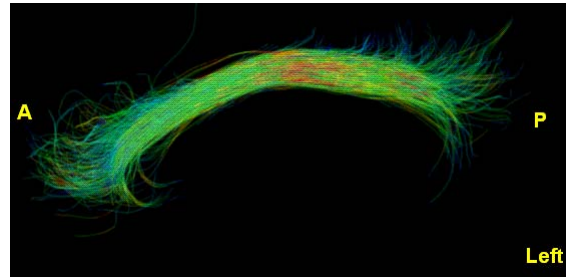
Figure 6-20 shows the left-right asymmetry of the fractional anisotropy in the superior part of the cingulum in control and schizophrenia groups. In both cases the FA is greater in the left cingulum and especially in the middle-toward-anterior part of the bundle. This observation is consistent with earlier findings reported for healthy subjects [68]. Another interesting observation is that the schizophrenia group exhibits a stronger asymmetry between the left and right cingulum.

Finally, Figure 6-21 shows the 3-D visualization of the p-values associated with the control vs. schizophrenia analysis of the FA values. The radius of the spheres at each sample of the cluster center represents the significance of the group differences.

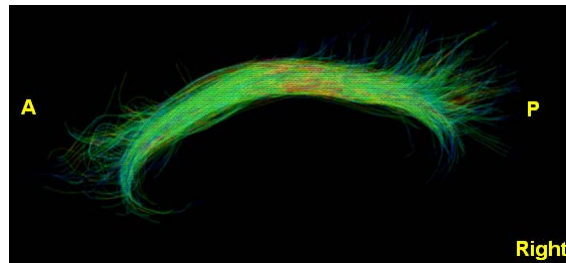
6.3.3 Discussion

As shown in the previous sections, our tract-oriented method is able to reveal the local variations of the diffusion parameters. This approach is thus very sensitive compared to ROI-based methods, where the differences might be lost when parameters (such as FA) are being averaged over the entire tract and compared between groups. Although voxel-based methods are potentially able to show the local variations, the results are sensitive to registration errors.

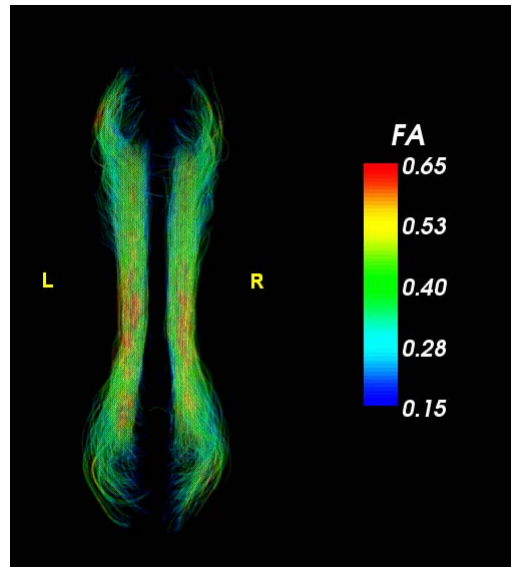
Abnormalities within the genu of the corpus callosum are localized in the midsagittal region, thus appear in the portion of the tract where the fibers of the tract run in isolation, which suggests disruptions in tracts interconnecting frontal lobes. On the other hand, changes observed in the schizophrenia group in the parasagittal, but not midsagittal portion of the splenium of the corpus callosum are suggestive of factors other than disruption of corpus callosum integrity, such as interfering and



(a)



(b)



(c)

Figure 6-17: Clustered trajectories of (a) left and (b) right cingulum from 44 subjects are shown in the atlas space. (c) shows these bundles from the axial view which each voxel is colored based on its FA value.

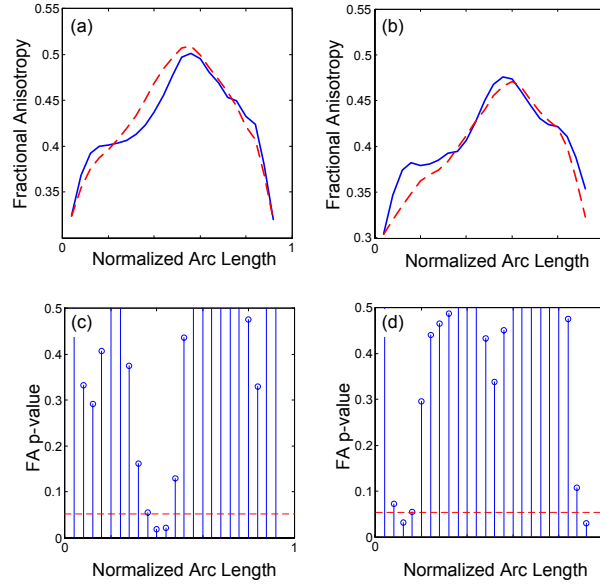


Figure 6-18: Comparison of the fractional anisotropy along the arc length of the superior part of the (a) left and (b) right cingulum in control (solid blue) and schizophrenia (dashed red) cases and (c), (d) the corresponding p-values. Increased FA is observed in the middle of left cingulum, while decreased FA is seen in the anterior part of the right cingulum.

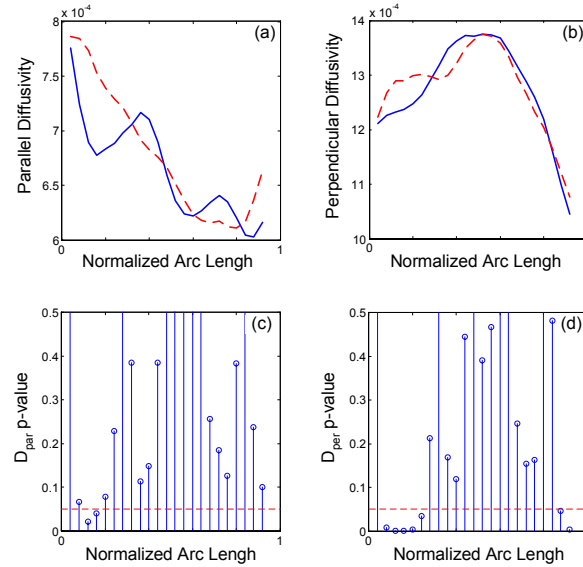


Figure 6-19: Comparison of the (a) parallel and (b) perpendicular diffusivity along the arc length of the superior part of the right cingulum in control (solid blue) and schizophrenia (dashed red) cases and (c), (d) the corresponding p-values. The anterior part of the bundle shows increased diffusivity in all directions. Hence the FA analysis in Figure 6-18(b) does not completely reflect the changes in the right cingulum.

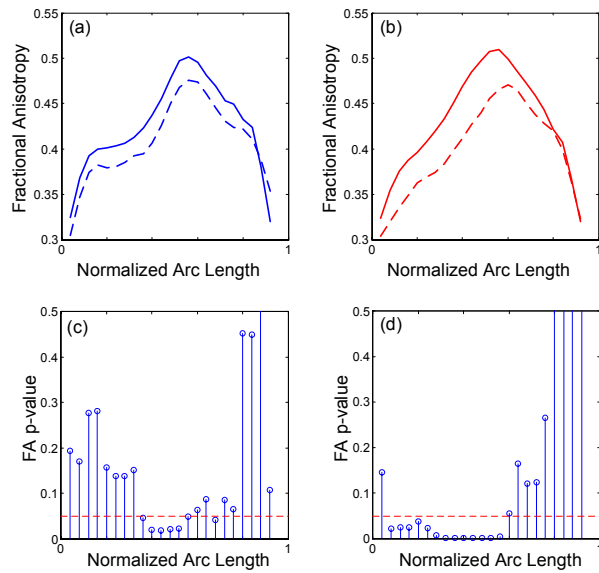


Figure 6-20: Comparison of the fractional anisotropy along the normalized arc length of the superior part of the left (solid lines) and right (dashed lines) cingulum and for the (a) control and (b) schizophrenia cases. The corresponding p-values in (c) and (d) demonstrates statistically significant left-greater-than-right asymmetry in both groups. A more important observation is that the asymmetry is more pronounced in the schizophrenic cases.

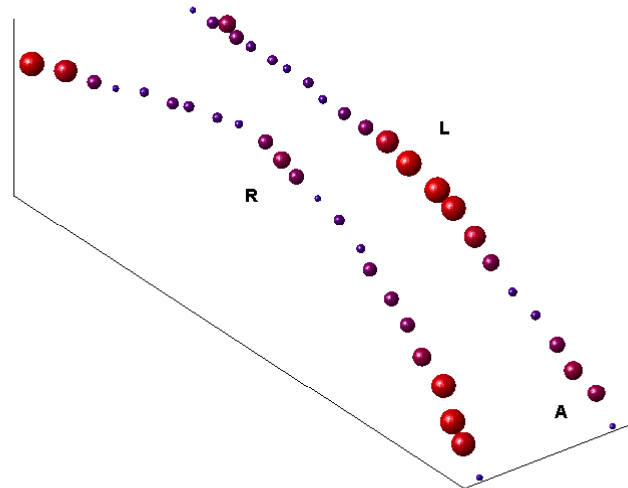


Figure 6-21: Three-dimensional visualization of the significance of the group analysis performed for the FA values between the control and schizophrenia subjects.

crossing fibers.

We would like to stress that a comprehensive DTI analysis requires examination of both anisotropy measures, such as FA, and the diffusivity magnitude, e.g., MD, or equivalently the parallel and perpendicular diffusivity eigenvalues [40]. The diffusivity perpendicular to the fiber tract reflects changes in axonal membrane, myelin, or extracellular space [71], whereas membrane disintegration and gliosis may create diffusion barriers in the direction of the fiber tract and reduce the parallel diffusivity [71]. Axonal damage may also result in reduced diffusivity perpendicular to the fiber tract [71]. Further studies, however, are still needed to understand better the relationship between micropathology and diffusion measures. Specifically, the relationship between diffusion measures and brain anatomy, including influence of other anatomical structures on anisotropy measures (as demonstrated on splenium as well as UF), need to be better understood. The analysis performed on the uncinate fasciculus clearly demonstrates the fact that fractional anisotropy alone is not sufficient to study white matter integrity.

Another important issue, which is essential in group analysis and was discussed in previous schizophrenia studies [46], is the age and/or disease-dependence of the diffusion parameters. To illustrate this point, in Figure 6-22 we have plotted the age-dependence of the fractional anisotropy in the mid-sagittal point of the lower splenium. While healthy subjects exhibit small age-dependence in the age range studied, considerable reduction of FA is observed in schizophrenic cases upon aging.

6.4 Summary

In this chapter, we first showed that with our clustering approach, the fiber bundles can be modeled by their mean trajectory and their spatial variation and can be efficiently visualized in 3-D. Preliminary results on a brain development study enabled by our tract-oriented quantitative analysis were also presented. We showed that the changes in FA in a developing brain reveal themselves to be more significant locally rather than when averaged over the whole ROI.

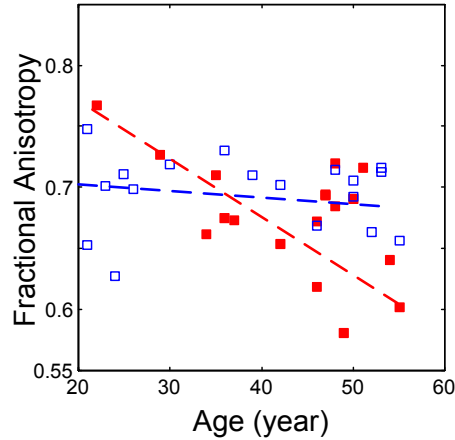


Figure 6-22: Age-dependence of the fractional anisotropy in mid-sagittal point of the lower splenium for healthy (open squares) and schizophrenic (closed squares) subjects.

A tract-oriented quantitative analysis of genu, splenium, and left and right uncinate fasciculus, inferior longitudinal fasciculus, inferior fronto-occipito fasciculus and cingulum to identify group differences in the schizophrenia group was presented. The method is able to reveal the local variations of the fiber integrity, which are lost when the quantitative parameters are averaged over the entire fiber tract in ROI-based methods. Significant reduction in fractional anisotropy in the left part of the genu and bilaterally in the middle of each side of the splenium was observed. The analysis of uncinate fasciculus showed a statistically significant increase in FA and a trend of increase in parallel diffusivity in the middle part of uncinate fasciculus as well as a pattern of asymmetry between left and right. We also observed significant increase in the perpendicular diffusivity for a large portion of the right inferior occipito-fronto fasciculus. The analysis on the left and right cingulum bundles revealed an increased FA in the middle of left cingulum, and decreased FA in the anterior part of the right cingulum. It also showed statistically significant left-greater-than-right asymmetry in both groups while this asymmetry was more pronounced in the schizophrenic cases. Finally, it was shown that age-dependence of the diffusivity parameters is an important factor when performing group analysis. Our findings presented in this chapter demonstrated the strength of tract-oriented quantitative analysis in revealing the group differences.

Chapter 7

Conclusions

In this dissertation, we have proposed and implemented a novel framework for robust clustering and quantitative analysis of cerebral white matter structures from diffusion MRI data. Our method automatically labels the trajectories of white matter pathways, extracted by a tractography algorithm, to anatomically known bundles of white matter fiber tracts. As has been shown in this thesis, such label assignment is done probabilistically. This approach addresses the uncertainty that exists in the input data and provides a principled way for us to identify the outliers and control the spread of the bundles.

By establishing point-correspondences between the trajectories in a cluster, our method has allowed us to perform statistical analysis along those trajectories and hence enabled us to quantify the local changes in diffusion properties of white matter along the bundles of fiber tracts. The estimation of the mean trajectory and the covariance matrix for each bundle has also let us construct three-dimensional abstract models that are beneficial in neurosurgical applications.

In this thesis, we developed a mixture model clustering to group fiber trajectories that are similar in shape and adjacent in spatial coordinates. To our knowledge, such a technique had not been used in this context before. We performed the mixture modeling on the *distances* between the trajectories as opposed to the coordinates of trajectories, and demonstrated how this technique performs well on variable-length feature vectors. We also demonstrated that a gamma mixture model accurately rep-

resents the nature of the distance of the 3-D trajectories from the cluster centers. We inferred the unknown parameters of the models in an Expectation Maximization setup. We also introduced a robust and computationally efficient approach for calculating the similarity and point correspondence between the trajectories, by building a distance map for each cluster center and a corresponding Voronoi diagram of labels.

We showed that the proposed framework has the flexibility to benefit from spatial anatomical knowledge, provided by an atlas of white matter fiber tracts, as prior information. We presented two Bayesian approaches that each incorporates anatomical information in the clustering of the fiber trajectories, at different levels of the Bayesian network. Integration of anatomical knowledge fully automates the process of quantitative analysis from specifying the seed points for tractography, guiding the clustering, to the labeling of the fiber tracts to anatomically known bundles. As shown in this thesis, it also makes the results more robust and less sensitive to the parameter settings.

To the best of our knowledge, our method represents the first implementation in which an anatomical prior is used in a mathematically principled framework and it is the first to identify group differences in diffusion properties of brain structures between healthy and pathological cases along the bundles of fiber tracts. The data analysis, especially our population study and findings on schizophrenia, demonstrated the capability, efficiency, and importance of the presented method in this thesis for clinical applications.

Appendix A

Graphical Models

Graphical models are a powerful modeling tool for dealing with the complexity and uncertainty of systems and are playing an important role in design and analysis of machine learning algorithms. They provide an intuitive representation of the system as a combination of smaller parts, where probability theory ensures the mathematical consistency of the model and the proper interaction of the model and data.

Probabilistic graphical models are graphs in which random variables are represented by nodes connected by edges to represent conditional dependencies. In fact, conditional dependence is the key to efficient probabilistic representation and computation and is expressed very clearly in graphical models. So, in addition to the graph structure, the conditional probability distribution (CPD) at each node, conditioned on its parents, must be specified. For discrete variables this can be shown by a table (CPT), which lists the probability that the child node takes on each of its different values for each combination of values of its parents.

Major types of graphical models include: directed, undirected (also called Markov Random Fields (MRFs) or Markov Networks), and factor graphical models.

In this work we deal with directed graphical models, also known as Bayesian Networks or Belief Networks (BNs). An example is shown in Figure A-1. The model gives a factorization of the joint probability distribution over variables as a product

of the probability distribution of each variable conditioned on its parents, i.e.,

$$p(x_1, \dots, x_k) = \prod_{i=1}^k p(x_i | Pa(X_i)), \quad (\text{A.1})$$

where $Pa(X_i)$ is the set of parents of X_i . Major independences are:

- Markov Property: Every node is conditionally independent of its non-descendants, given its parents.
- Each node is unconditionally (a priori) independent of any non-descendant non-ancestor, unless they share a common ancestor.
- Each node is conditionally independent of all others given its Markov blanket, i.e. set of its parents, children, and children's parents.

The conditional independence are best explained by means of the “Bayes Ball” algorithm: Two (sets of) nodes A and B are conditionally independent (d-separated) given a set C if and only if there is no way for a ball to get from A to B in the graph. Before explaining the allowable movements, a distinction should be made between hidden and observed nodes. Hidden nodes are nodes whose values are not known, and are depicted as unshaded, while observed nodes are shaded.

Figure A-1 shows different situations where nodes X_1 and X_3 are either dependent or independent, based on their connection to X_2 . In the first column X_2 is a “leaf”, i.e., a node with two parents. If X_2 is hidden, its parents are marginally independent, and hence the ball does not pass through. However, if X_2 is observed, the parents become dependent because of the explaining away phenomenon, i.e., the parents “compete” to explain the observed node.

In the second column, X_2 is a “root”, i.e., two edges diverge from X_2 . If X_2 is hidden, the children are dependent, because they have a hidden common cause. If X_2 is observed, its children are conditionally independent,

Finally, the third column shows a case, in which X_2 has one incoming and one outgoing arrow. The nodes upstream and downstream of X are dependent if and only if X_2 is hidden, because if X_2 is observed, conditioning on this node breaks the graph.

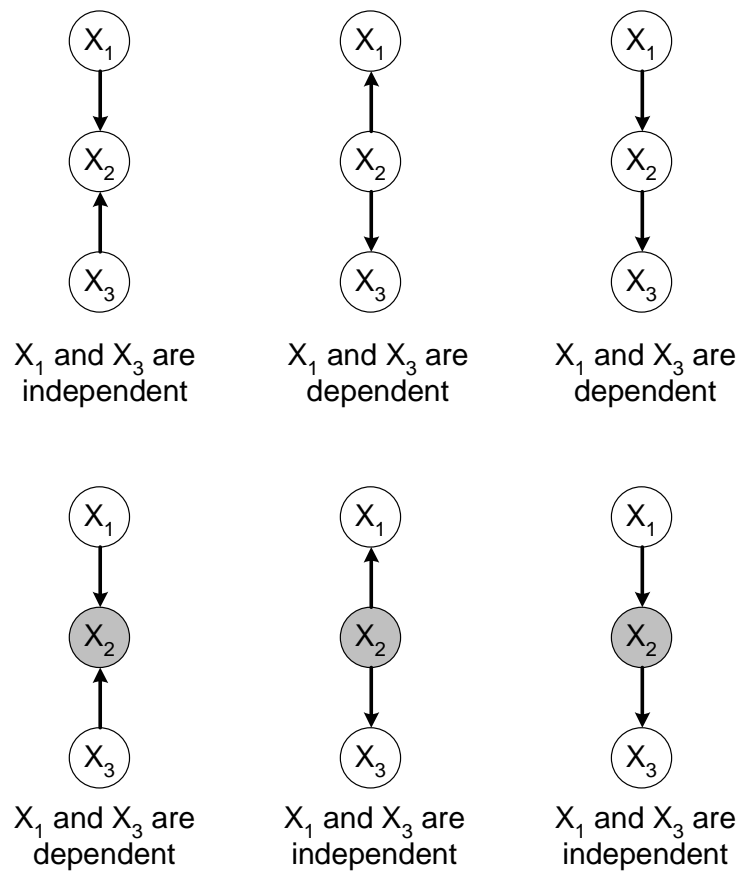


Figure A-1: Different situations where nodes X_1 and X_3 are either dependent or independent, based on their connection to X_2 .

Bibliography

- [1] I Agartz, J.L.R. Andersson, and S. Skare. Abnormal white matter in schizophrenia: a diffusion tensor imaging study. *Neuroreport*, 12:2251–2254, 2001.
- [2] D. C. Alexander, C. Pierpaoli, P. J. Basser, and J. C. Gee. Spatial transformations of diffusion tensor magnetic resonance images. *IEEE Trans. Med. Imag.*, 20:1131–1139, 2001.
- [3] H. Alt and M. Godeau. Computing the fréchet distance between two polygonal curves. *Int. J. Computational Geometry Applications*, pages 75–79, 1995.
- [4] B. Ardekani, J. Nierenberg, M.J. Hoptman, D.C. Javitt, and K.O. Lim. MRI study of white matter diffusion anisotropy in schizophrenia. *Neuroreport*, 14:2025–2029, 2003.
- [5] E. Arkin, P. Chew, D. Huttenlocher, K. Kedem, and J. Mitchel. An efficiently computable metric for comparing polygonal shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13:209–215, 1991.
- [6] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Mag. Res. Med.*, 56:411–421, 2006.
- [7] P. Basser. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR in Biomedicine*, 8:333344, 1995.
- [8] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44:625–632, 2000.

- [9] P.J. Basser, J. Mattiello, and D. LeBihan. Estimation of the effective self-diffusion tensor from the nmr spin echo. *J. Magnetic Resonance*, 103:47–54, 1994.
- [10] P.J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.*, 44:625–632, 2000.
- [11] P.G. Batchelor, F. Calamante, J. D. Tournier, D. Atkinson, D. L. Hill, and A. Connelly. Quantification of the shape of fiber tracts. *Magn. Reson. Med.*, 55:894–903, 2006.
- [12] C. Beaulieu. The basis of anisotropic water diffusion in the nervous system – a technical review. *NMR Biomed.*, 15:435–455, 2001.
- [13] J. I. Berman, P. Mukherjee, S. C. Partridge, S. P. Miller, D. M. Ferriero, A. J. Barkovich, D. B. Vigneron, and R. G. Henry. Quantitative diffusion tensor MRI fiber tractography of sensorimotor white matter development in premature infants. *NeuroImage*, 27:862–871, 2005.
- [14] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2006.
- [15] M. Bozzali, A. Falini, M. Franceschi, M. Cercignani, M. Zuffi, G. Scotti, G. Comi, and M. Filippi. White matter damage in Alzheimer’s disease assessed in vivo using diffusion tensor magnetic resonance imaging. *J. Neurol. Neurosurg. Psychiatry*, 72:742–746, 2002.
- [16] A. Brun, M. Bjornemo, R. Kikinis, and C.-F. Westin. White matter tractography using sequential importance sampling. In *ISMRM*, page 1131, 2003.
- [17] A. Brun, H. Knutsson, H. J. Park, M. E. Shenton, and C.-F. Westin. Clustering fiber tracts using normalized cuts. In *Seventh International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI’04)*, Lecture Notes in Computer Science, pages 368–375, Rennes - Saint Malo, France, September 2004.

- [18] M.S. Buchsbaum, J. Friedman, B.R. Buchsbaum, K.W. Chu, E.A. Hazlett, R. Newmark, J.S. Schneiderman, Y. Torosjan, C. Tang, P.R. Hof, D. Stewart, K.L. Davis, and J. Gorman. Diffusion tensor imaging in schizophrenia. *Biological Psychiatry*, 60:1181–1187, 2006.
- [19] J. Burns, D. Job, M.E. Bastin, H. Whalley, T. Macgillivray, E.C. Johnstone, and S.M. Lawrie. Structural disconnectivity in schizophrenia: a diffusion tensor magnetic resonance imaging study. *Br. J. Psychiatry*, 182:439–443, 2003.
- [20] Y. Cao, M. I. Miller, R. L. Winslow, and L. Younes. Large deformation diffeomorphic metric mapping of vector fields. *IEEE Trans. Medical Imaging*, 24:1216–1230, 2005.
- [21] M.-C. Chiang, A. D. Leow, A. D. Klunder, R. A. Dutton, M. Barysheva, S. E. Rose, K. L. McMahon, G. I. de Zubicaray, A. W. Toga, and P. M. Thompson. Fluid registration of diffusion tensor images using information theory. *IEEE Trans. Medical Imaging*, 27:442–456, 2008.
- [22] I. Corouge, P. T. Fletcher, S. Joshi, S. Gouttard, and G. Gerig. Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. *Med. Image Anal.*, 10:786–798, 2006.
- [23] I. Corouge, S. Gouttard, and G. Gerig. Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In *IEEE Int. Symp. Biomed. Imag.*, pages 344–347, 2004.
- [24] G. Cortelazzo, G. A. Mian, G. Vezzi, and P. Zamperoni. Trademark shapes description by string-matching techniques. *Pattern Recognition*, 27:1005–1018, 1994.
- [25] Le Bihan. D., J. F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: concepts and applications. *J. Magn. Reson. Imaging*, 13:534–546, 2001.

- [26] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Trans. Medical Imaging*, 21:525–537, 2002.
- [27] K.L. Davis, D.G. Stewart, J.I. Friedman, M. Buchsbaum, P.D. Harvey, P.R. Hof, J. Buxbaum, and V. Haroutunian. White matter changes in schizophrenia: Evidence for myelin-related dysfunction. *Arch. Gen. Psychiatry*, 60:443–456, 2003.
- [28] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39:138, 1977.
- [29] R. Deriche and O. Faugeras. 2-d curve matching using high curvature points: application to stereo vision. In *Proc. Int. Conf. Pattern Recognition*, pages 240–242, 1990.
- [30] Z. Ding, J. C. Gore, and A. W. Anderson. Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. *Mag. Reson. Med.*, 49:716–721, 2003.
- [31] Zhaohua Ding, John C. Gore, and Adam W. Anderson. Case study: reconstruction, visualization and quantification of neuronal fiber pathways. In *VIS '01: Proceedings of the conference on Visualization '01*, pages 453–456, 2001.
- [32] J. Dubois, L. Hertz-Pannier, G. Dehaene-Lambertz, Y. Cointepas, and D. Le Bihan. 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*, 30:1121–1132, 2006.
- [33] Frank Enders, Natascha Sauber, Dorit Merhof, Peter Hastreiter, Christopher Nimsky, and Marc Stamminger. Visualization of white matter tracts with wrapped streamlines. In *IEEE Visualization 2005 (VIS 2005)*, pages 51–58, 2005.

- [34] A. Fick. On liquid diffusion. *Philosophical Mag. and J. Science*, 10:30–39, 1855.
- [35] P. T. Fletcher and S. Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87:250–262, 2007.
- [36] J. Foong, M. Maier, C. A. Clark, G. J. Barker, D. H. Miller, and M. A. Ron. Neuropathological abnormalities of the corpus callosum in schizophrenia: a diffusion tensor imaging study. *J. Neurol. Neurosurg. Psychiatry*, 68:242–244, 2000.
- [37] J. Foong, M.R. Symms, G.J. Barker, M. Maier, and D.H. Miller. MRI study of white matter diffusion anisotropy in schizophrenia. *Neuroreport*, 13:333–336, 2002.
- [38] Ola Friman, Gunnar Farneback, and Carl-Fredrik Westin. A Bayesian approach for stochastic white matter tractography. *TMI*, 25(8):965–978, 2006.
- [39] G. Gerig, S. Gouttard, and I. Corouge. Analysis of brain white matter via fiber tract modeling. In *Proc. IEEE Int. Conf. EMBS*, pages 4421–4424, 2004.
- [40] K.M Hasan. Diffusion tensor eigenvalues or both mean diffusivity and fractional anisotropy are required in quantitative clinical diffusion tensor mr reports: Fractional anisotropy alone is not sufficient. *Radiology*, 239:611–613, 2006.
- [41] K. Hua, J. Zhang, S. Wakana, H. Jiang, X. Li, D. S. Reich, P. A. Calabresi, J. J. Pekar P. C.M. van Zijl, and S. Mori. Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. *NeuroImage*, 39:336–347, 2008. to be published.
- [42] H. Huang, J. Zhang, S. Wakana, W. Zhang, T. Ren, L. J. Richards, P. Yarowsky, P. Donohue, E. Graham, P. C.M. v. Zijl, and S. Mori. White and gray matter development in human fetal, newborn and pediatric brains. *Neuroimage*, 33:27–38, 2006.

- [43] P. Indyk and S. Venkatasubramanian. Approximate congruence in nearly linear time. *Computational Geometry*, 24:115–128, 2002.
- [44] B. J. Jellison, A. S. Field, J. Medow, M. Lazar, M. S. Salamat, and A. L. Alexander. Diffusion tensor imaging of cerebral white matter: A pictorial review of physics, fiber tract anatomy, and tumor imaging patterns. *American Journal of Neuroradiology*, 25:356–369, 2004.
- [45] L. Jonasson, X. Bresson, P. Hagmann, O. Cuisenaire, R. Meuli, and J.-P. Thiran. White matter fiber tract segmentation in dt-mri using geometric flows. *Medical Image Analysis*, 9:223–236, 2005.
- [46] D.K. Jones, M. Catani, C. Pierpaoli, S.J. Reeves, S.S. Shergill, M. O’Sullivan, P. Golesworthy, P. McGuire, M.A. Horsfield, A. Simmons, S.C. Williams, and R.J. Howard. Age effects on diffusion tensor magnetic resonance imaging tractography measures of frontal cortex connections in schizophrenia. *Human Brain Mapping*, 27:230–238, 2006.
- [47] R.A. Kanaan, S.S. Shergill, G.J. Barker, M. Catani, V.W. Ng, R. Howard, P.K. McGuire, and D.K. Jones. Tract-specific anisotropy measurements in diffusion tensor imaging. *Psychiatry Res.*, 146(1):73–82, 2006.
- [48] Gordon Kindlmann, Raul San Jose Estepar, Marc Niethammer, Steven Haker, and Carl-Fredrik Westin. Geodesic-loxodromes for diffusion tensor interpolation and difference measurement. In *Tenth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI’07)*, pages 1–9, October 2007.
- [49] M. Kubicki, C.-F. Westin, S.E. Maier, H. Mamata, M. Frumin, H. Ernst-Hirschfeld, R. Kikinis, F.A. Jolesz, R.W. McCarley, and M.E. Shenton. Cingulate fasciculus integrity disruption in schizophrenia: a magnetic resonance diffusion tensor imaging study. *Biological Psychiatry*, 54:1171–1180, 2003.

- [50] B. J. Laundre, B. J. Jellison, B. Badie, Alexander A. L., and A. S. Field. Diffusion tensor imaging of the corticospinal tract before and after mass resection as correlated with clinical motor findings: Preliminary data. *American Journal of Neuroradiology*, 26:791–796, 2005.
- [51] M. Lazar, D. M. Weinstein, J. S. Tsuruda, . M. Hasan, K. Arfanakis, M. E. Meyerand, B. Badie, H. A. Rowley, V. Haughton, A. Field, and A. L. Alexander. White matter tractography using diffusion tensor deflection. *Human Brain Mapping*, 18:306–321, 2003.
- [52] A. Leemans, J. Sijbers, S. De Backer, E. Vandervliet, and P. Parizel. Multiscale white matter fiber tract coregistration: A new feature-based approach to align diffusion tensor data. *Mag. Res. Med.*, 55:1414–1423, 2006.
- [53] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor mri processing. *J. Math. Imaging Vis.*, 25:423444, 2006.
- [54] C. Lenglet, M. Rousson, R. Deriche, O. Faugeras, S. Lehericy, and K. Ugurbil. A riemannian approach to diffusion tensor images segmentation. In *Information Processing in Medical Imaging*, pages 591–602, 2005.
- [55] Catani M., Howard R.J., Pajevic S., and Jones D.K. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *NeuroImage*, 17:77–94(18), 2002.
- [56] M. Maddah, W. E. L. Grimson, and S. K. Warfield. Statistical modeling and EM clustering of white matter fiber tracts. In *IEEE Int. Symp. Biomedical Imaging*, pages 53–56, 2006.
- [57] M. Maddah, W.E.L. Grimson, S.K. Warfield, and W.M. Wells. A unified framework for clustering and quantitative analysis of white matter fiber tracts. *Med. Image Anal.*, 12:191–202, 2008.

- [58] M. Maddah, A. Mewes, S. Haker, W. E. L. Grimson, and S. K. Warfield. Automated atlas-based clustering of white matter fiber tracts from DTMRI. In *MICCAI*, pages 188–195, 2005.
- [59] M. Maddah, W. M. Wells, S. K. Warfield, C-F Westin, and W. E. L. Grimson. Probabilistic clustering and quantitative analysis of white matter fiber tracts. In *IPMI*, pages 372–383, 2007.
- [60] M. Maddah, L. Zöllei, C.-F. Westin, W. E. L. Grimson, and W. M. Wells. A mathematical framework for incorporating anatomical knowledge in DT-MRI analysis. In *IEEE Int. Symp. Biomedical Imaging*, 2008.
- [61] Mahnaz Maddah, Lilla Zöllei, W. Eric L. Grimson, and William M. Wells. Modeling of anatomical information in clustering of white matter fiber trajectories using dirichlet distribution. In *Mathematical Methods in Biomedical Image Analysis (MMBIA)*, Alaska, US, 2008.
- [62] Y. Masutani, S. Aoki, O. Abe, and K. Ohtomo. Model-based tractography based on statistical atlas of MR-DTI. In *Int. Symp. Biomedical Imaging*, pages 89–92, 2006.
- [63] Lauren O’Donnell. *Cerebral White Matter Analysis Using Diffusion Imaging*. Harvard-MIT Division of Health Sciences and Technology PhD Thesis.
- [64] L. O’Donnell and C.-F. Westin. White matter tract clustering and correspondence in populations. In *Eighth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI’05)*, Lecture Notes in Computer Science 3749, pages 140–147, Palm Springs, CA, USA, October 2005.
- [65] Lauren J. O’Donnell and Carl-Fredrik Westin. Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Trans. Medical Imaging*, 26:1562–1575, 2007.

- [66] Lauren J. O'Donnell, Carl-Fredrik Westin, and Alexandra J. Golby. Tract-based morphometry. In *Int. Conf. Medical Image Computing and Computer-Assisted Intervention*, volume 4792, pages 161–168, 2007.
- [67] H. J. Park, M. Kubicki, M. E. Shenton, A. Guimond, R. W. McCarley, S.E. Maier, R. Kikinis, F. A. Jolesz, and C. F. Westin. Spatial normalization of diffusion tensor MRI using multiple channels. *NeuroImage*, 20:1995–2009, 2003.
- [68] H. J. Park, C.-F. Westin, M. Kubicki, S. E. Maier, M. Niznikiewicz, A. Baer, M. Frumin, R. Kikinis, F. A. Jolesz, R. W. McCarley, and M. E. Shenton. White matter hemisphere asymmetries in healthy subjects and in schizophrenia: A diffusion tensor MRI study. *NeuroImage*, 24:213–223, 2004.
- [69] S. C. Partridge, P. Mukherjee, J. I. Berman, R. G. Henry, S. P. Miller, Y. Lu, O. A. Glenn, D. M. Ferriero, A. J. Barkovich, and D. B. Vigneron. Tractography-based quantitation of diffusion tensor imaging parameters in white matter tracts of preterm newborns. *J Magn Reson Imaging*, 22:467–474, 2005.
- [70] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *Int. J. Comp. Vis.*, 66:41–66, 2006.
- [71] C. Pierpaoli and P.J. Basser. Toward a quantitative assessment of diffusion anisotropy. *Magn. Reson. Med.*, 36:893–906, 1996.
- [72] M. Pollari, T. Neuvonen, and Lötjönen. Affine registration of diffusion tensor MR images. In *MICCAI*, pages 629–636, 2006.
- [73] D. Prayer, A. J. Barkovich, D. A. Kirschner, L. M. Prayer, T. P. L. Roberts, J. Kucharczyk, and M. E. Moseley. Visualization of nonstructural changes in early white matter development on diffusion-weighted mr images: Evidence supporting premyelination anisotropy. *American J. Neuroradiology*, 22:1572–1576, 2001.

- [74] G. K. Rohde, S. Pajevic, and C. Pierpaoli. Multi-channel registration of diffusion tensor images using directional information. In *ISBI*, pages 712–715, 2004.
- [75] G. K. Rohdem, S. Pajevic, C. Pierpaoli, and P. J. Basser. A comprehensive approach for multi-channel image registration. In *Proceedings of 2nd International Workshop on Biomedical Image Registration*, page 214223, Philadelphia, PE, 2003.
- [76] M. Rousson, C. Lenglet, and R. Deriche. Level set and region based surface propagation for diffusion tensor mri segmentation. In *MICCAI*, pages 123–134, 2004.
- [77] J. Ruiz-Alzola, C.-F. Westin, S. K. Warfield, A. Nabavi, and R. Kikinis. Non-rigid registration of 3d scalar vector and tensor medical data. In *Third International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 541–550, Pittsburgh, October 11–14 2000.
- [78] D.H Salat, D.S. Tuch, D.N. Greve, A.J.W van der Kouwe, N.D. Hevelone, A.K. Zaleta, B.R. Rosen, B. Fischl, S. Corkin, H.D. Rosas, and A.M. Dale. Age-related alterations in white matter microstructure measured by diffusion tensor imaging. *Neurobiology of Aging*, 26:1215–1227, 2005.
- [79] M.F. Schocke, K. Seppi, R. Esterhammer, C. Kremser, K.J. Mair, B.V. Czermak, W. Jaschke, W. Poewe, and G.K. Wenning. Trace of diffusion tensor differentiates the parkinson variant of multiple system atrophy and parkinson’s disease. *NeuroImage*, 21:14431451, 2004.
- [80] T. Schonberg, P. Pianka, T. Hendler, and Y. Assaf. Characterization of displaced white matter tracts using DTI and fMRI. In *Proc. ISMRM*, page 468, 2005.

- [81] J. S. Shimony, A. Z. Snyder, N. Lori, and T. E. Conturo. Automated fuzzy clustering of neuronal pathways in diffusion tensor tracking. In *Proc. Int. Soc. Mag. Reson. Med.*, page 10, 2002.
- [82] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Z. Cader, P. M. Matthews, and T. E. J. Behrens. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31:1487–1505, 2006.
- [83] S.-K. Song, S.-W. Sun, M. J. Ramsbottom, C. Chang, J. Russell, and A. H. Cross. Dysmyelination revealed through mri as increased radial (but unchanged axial) diffusion of water. *NeuroImage*, 17:1429–1436, 2002.
- [84] R.M. Steel, M.E. Bastin, S. McConnell, I. Marshall, D.G. Cunningham-Owens, S.M. Lawrie, E.C. Johnstone, and J.J.K. Best. Diffusion tensor imaging (DTI) and proton magnetic resonance spectroscopy (1H MRS) in schizophrenic subjects and controls. *Psych. Res.: Neuroimage*, 106:161–170, 2001.
- [85] E. O. Stejskal and J. E. Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J. Chem. Phys.*, 42:288–292, 1965.
- [86] Z. Sun, F. Wang, L. Cui, J. Breeze, X. Du, X. Wang, Z. Cong, H. Zhang, B. Li, N. Hong, and D. Zhang. Abnormal anterior cingulum in patients with schizophrenia: a diffusion tensor imaging study. *Neuroreport*, 14:1933–1836, 2003.
- [87] I. F. Talos, L. O’Donnell, C.-F. Westin, S. K. Warfield, W. M. Wells, S. S. Yoo, L. Panych, A. Golby, H. Mamata, S. E. Maier, P. Ratiu, C. G. Guttman, P. Mc. L. Black, F. A. Jolesz, and R. Kikinis. Diffusion tensor and functional MRI fusion with anatomical mri for image guided neurosurgery. In *Sixth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI’03)*, pages 407–415, 2003.

- [88] A. Tsai, C.-F. Westin, A. O. Hero, and A. S. Willsky. Fiber tract clustering on manifolds with dual rooted-graphs. In *IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
- [89] D. S. Tuch. Q-ball imaging. *Magnetic Resonance in Medicine*, 52:1358 – 1372, 2004.
- [90] D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine*, 48:577–582, 2002.
- [91] R. C. Veltkamp. Shape matching: Similarity measures and algorithms. In *Int. Conf. Shape Modeling and Applications*, pages 188–197, 2001.
- [92] S. Wakana, H. Jiang, L. M. Nagae-Poetscher, P. C. M. van Zijl, and S. Mori. Fiber tract-based atlas of human white matter anatomy. *Radiology*, 230:77–87, 2004.
- [93] Z. Wang and B. C. Vemuri. Dti segmentation using an information theoretic tensor dissimilarity measure. *IEEE Trans. Medical Imaging*, 24:1267– 1277, 2005.
- [94] S.K. Warfield, K.H. Zou, and W.M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Medical Imaging*, 23:903–921, 2004.
- [95] W.M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1:35–51, 1996.
- [96] D. J. Werring, C. A. Clark, G. J. Barker, A. J. Thompson, and D. H. Miller. Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology*, 52:1626–1632, 1999.

- [97] C.-F. Westin, S. E. Maier, H. Mamata, A. Nabavi, F. A. Jolesz, and R. Kikinis. Processing and visualization of diffusion tensor MRI. *Medical Image Analysis*, 6(2):93–108, 2002.
- [98] T. H. Williams, N. Gluhbegovic, and J. Y. Jew. The human brain: Dissections of the real brain. virtual hospital. The University of Iowa, College of Medicine, 1997. <http://www.vh.org/adult/provider/anatomy/BrainAnatomy/BrainAnatomy.html>.
- [99] P. A. Yushkevich, H. Zhang, T. J. Simon, and J. C. Gee. Structure-specific statistical mapping of white matter tracts using the continuous medial representation. In *IEEE Int. Conf. Computer Vision*, 2007.
- [100] H. Zhang, P. A. Yushkevich, D. C. Alexander, and J. C. Gee. Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Medical Image Analysis*, 10(5):764–785, 2006.
- [101] U. Ziayn, M. R. Sabuncu, W. E. L. Grimson, and C.-F. Westin. A robust algorithm for fiber-bundle atlas construction. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 2007.
- [102] U. Ziyen, M. R. Sabuncu, L. J. O’Donnell, and C.-F. Westin. Nonlinear registration of diffusion mr images based on fiber bundles. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI ’07)*, volume 4791 of *Lecture Notes in Computer Science*, pages 351–358, 2007.
- [103] L. Zöllei, E. Learned-Miller, E. Grimson, and W.M. Wells. Efficient population registration of 3D data. In *Workshop on Computer Vision for Biomedical Image Applications: Current Techniques and Future Trends*, 2005.
- [104] Lilla Zöllei. *A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images*. 2006. Ph.D. thesis, MIT.