Multi-subject Connectivity-Based Parcellation

Erzhuo Wang

A thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science

University of Washington

2013

Reading Committee:

David Haynor

Colin Studholme

Paul Kinahan

Program Authorized to Offer Degree:

Department of Bioengineering

ABSTRACT

Connectivity has been proposed as a criterion for functional-anatomic segregation of cortical areas. In this thesis, I present a new method of characterizing the DTI-based connectivity profile of cortical voxels using Gaussian mixture models (GMMs). A variety clustering techniques are applied to perform the connectivity-based parcellation (CBP). I first parcellated the human inferior parietal lobule (IPL) on connectivity profiles using spectral clustering and a hidden Markov random field (HMRF) model. I applied our approach to multi-subject parcellation. I then segmented other cortical areas such as precentral and postcentral cortex, using spectral non-parametric Bayes models. A new approach resolving crossing fibers with compressed sensing (CS) was also examined.

Using the multi-subject GMM-HMRF approach, results in a smoother segmentation of IPL that is independent of the set of subjects and visually consistent with the Juelich Atlas. The spectral non-parametric Bayes models enable data learn the number of segments automatically. The compressed sensing method is shown to significantly reduce the amount of data required and the computing time while preserving the accuracy.

Table of Contents

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors Dr. David Haynor and Dr. Thomas Grabowski for their support, guidance and their availability whenever I need an advice or help during my Master Thesis project. Thanks to all the IBIC Laboratory staffs for their help throughout my stay with them, they always have been very helpful in many ways; it was a great pleasure to work in this welcoming environment. Also special thanks to Rosalia Tungaraza who have been a good support to me in order to start my research and provide me some useful advice

I would also like to extend my appreciation to Dr. Colin Studholme for his help during my first rotation and for his trust in me to integrate the MSc. thesis. Thanks to all the academic staff of University of Washington. It has been a wonderful experience, in both academic and personal life.

Finally thanks to my family and friends for their support all along the two years, even far away.

Chapter 1 INTRODUCTION

The cortical field hypothesis states that human cerebral cortex is a mosaic of subregions, each with a specific structural and functional pattern (Zilles, et al., 2010). Brodmann's (1909) seminal work parcellated the cerebral cortex using cytoarchitectonics. Cellular microstructure, however, is only one structural feature, and brain function is dependent on both cytoarchitectonics and inter-region connectivity working together.

Diffusion tensor imaging (DTI) has made it possible to characterize and visualize the organization of white matter fiber traces in human. With the help of this modality in conjunction with probabilistic tractography (Basser, et al. 2000), one can decide whether an individual voxel of brain is connected to others. Recently, a number of studies have used diffusion weighted MRI to study the subdivision of cortex via connectional anatomy. This approach, connectivity-based parcellation (CBP), assumes that adjacent cortical voxels having similar connectivity patterns (tractogram) are part of one subregion (Johansen-Berg, et al, 2004). While the cytoarchitectonic work and DTI cannot be done in the same individuals, the results in selected brain areas, such as the inferior parietal lobule, below, are similar, suggesting that DTI may provide an approximate guide to cortical parcellation in vivo. The identification of connectional disparate sub-regions or human cortex is an important aspect of neurosciences and clinical surgery.

In the work reported here, we for the first time estimated tractograms with a Gaussian Mixture Model (GMM). A novel and efficient distance metric, the earth mover's distance was applied to the measurement of the similarity between connectivity patterns. Then a variety of clustering approaches were performed to segment the regions of interest (ROI). These techniques reveal some interesting features of the connectional anatomy of human cortex and proved to be efficient ways to practice CBP. A new method using compressed sensing to resolve crossing fibers of DWI images is also examined.

1.1 Background

In-vivo segregation of cerebral cortex into distinct functional sub-divisions is a significant task in neuroscience. This sub-division is usually referred to as "cortex parcellation". Under the premise that structure reflects function, understanding the structural organization of cortex is crucial for study brain functionality. Structural identification of functional imaging data has long been based on the cytoarchitectonic map developed by Brodmann in 1909. In recent years, connectivity-based parcellation (CBP) has become a popular approach to segmenting gray matter structures including thalamus and cerebral cortex, which exploits diffusion weighted MRI to reveal the distinct connectivity

patterns of each sub-division. CBP rests on the assumption that spatially contiguous voxels share common connectional profiles, which leads to common functionality (Anwander, et al, 2006).

Connectivity-based parcellation also has clinical interest. Abnormal connectivity patterns in cingulate subregions have been observed in diseases such as Alzheimer's disease, depression, schizophrenia, and obsessive compulsive disorder (Rosenberg, et al., 2004; Gotlib, et al., 2005; Kubicki, et al., 2005; Naggara, et al., 2006). Preterm infants have microstructural differences in cerebral white matter compared to term-born control infants in the absence of focal abnormalities (Anjari, et al., 2007).

Among the clinically significant regions of interest (ROI) are the inferior parietal lobule (IPL) and pre/post central gyri. IPL is involved in the perception of emotion in facial expressions, the interpretation of sensory information, spatially directed attention, phonological processing and other aspects of language (Radua, et al, 2010). The parcellation of IPL has been of great interest because of its functional and anatomical heterogeneity (Caspers, et al, 2011). The pre central and postcentral areas are highly related to the primary motor cortex and primary somatosensory cortex, respectively. The precentral gyrus is the main motor center planning and executing body movements; the postcentral gyrus is responsible for collecting and processing the sensory information from skin and from the spindle organs (Catani, et al., 2012). They have drawn particular attention from researchers for their interaction in neural connection.

Connectivity patterns were modeled as connectivity profiles (Anwander, et al, 2007) or fingerprints (Beckmann, et al, 2009), defined by the probabilities of connection between a cortical "seed" voxel and a set of predefined target regions. We have modified the existing CBP approaches by defining a connectivity pattern directly as a histogram of target voxel locations. A similarity matrix generated from such connectivity profiles is then segmented using clustering techniques such as spectral clustering (Shi & Malik, 2000), hidden Markov random fields (Zhang, et al, 2001) and non-parametric Bayes models (Murphy, 2012). Different clustering methods were applied to fit a variety of problems.

The modality we use in this research, diffusion tensor imaging (DTI), is limited in regions of crossing fibers. Most of the methods to resolve fiber crossing requiree large number of scan gradients and high b-values (Tuch, et al., 2002; Frank, et al., 2002; Tournier, et al., 2004), which is problematic for clinical time and hardware. We investigated a newly proposed approach (Landman, 2012), involving compressed sensing (Donoho, 2006). The number of gradients is expected to be highly reduced by employing this method.

1.2 Objectives

The main aspect of this thesis is to model tractograms with Gaussian mixture models, to compute the distance matrices between tractograms using a variety of metrics, and then to parcellate ROIs based on the distance matrices. Various clustering techniques are tested for parcellation: with a prior knowledge of the number of parcels (IPL), spectral clustering is used to solve the distance based clustering problem; for areas without a definite cluster number, such as pre/post central, we employ non-parametric Bayes approaches like Dirichlet process model to segment seed voxel and decide the proper number; hidden Markov random fields are mainly used to denoise the clustering obtained from the two methods mentioned.

Due to high heterogeneity of IPL, we also wished to develop a connectivity-based atlas built from multiple subjects. The CBP of the pre/post central area is focused on understanding the driving forces of parcellation in connectional structure.

I will also demonstrate the viability and efficiency of compressed sensing in resolving crossing fibers.

1.3 Organization

The structure of the thesis aims to explain the fundamental concept used in the connectivity modeling and segmentation of ROIs. The thesis is composed as follow:

- Chapter 1: Introduction, motivation and objectives of the project.
- Chapter 2: This chapter gives a general background of connectivity modeling, such as probabilistic tractography and Gaussian mixture models.
- Chapter 3: This chapter briefly introduce the clustering techniques used in this project, including spectral clustering, hidden Markov random fields and Dirichlet process model
- Chapter 4: A summary of Landman's method in resolution of crossing fibers is given in this chapter. I will describe the implementation and verification of this method.
- Chapter 5: Methodology
- Chapter 6: Results of experiments are presented and analysed.
- Chapter 7: Conclusion including the summary of the work done also the review of the future work is stated.

Chapter 2 MODELING OF CONNECTIVITY

Before entering into detailed connectivity-based parcellation, this chapter brings general background information about the diffusion tensor imaging, probabilistic tractography and the Gaussian mixture model of connectivity patterns.

2.1 Diffusion Tensor MRI

Diffusion tensor MRI (DTI) is the first noninvasive in vivo imaging modality which enables researchers to generate fiber-tract trajectories in soft fibrous tissues, such as nerves, muscles, ligaments, tendons, etc. (Basser, et al., 1994). Diffusion contrast is based on the self-diffusion of water molecules in tissue. (Le Bihan, et al., 1986). A diffusion-weighted pulse sequence is constructed by the addition of a pair of diffusion-sensitizing gradients, also known as motion-probing gradients, to a T2-weighted spin-echo sequence. This is known as Stejskal-Tanner diffusion encoding (Stejskal & Tanner, 1965). Molecular motion thus results in loss of signal intensity due to incomplete rephasing of water proton spins, which change position between and during the applications of the 2 diffusion-sensitizing gradients. This diffusion-weighted contrast can be fit to an exponential model (Stejskal-Tanner tensor model):

$$S_i = S_0 \cdot e^{-b \cdot ADC_i} \tag{2.1}$$

where S_i is the diffusion-weighted (DW) signal intensity observed at a given voxel with the diffusion-sensitizing gradients applied along direction *i*, and S_0 is the signal intensity at the same voxel measured without diffusion-sensitizing gradients. The apparent diffusion coefficient (ADC) is the diffusion constant measured in the clinical setting, reflecting the limitation that in vivo diffusion cannot be separated from other sources of water mobility, and *ADCi* is the *ADC* in the *i* direction. The *b* factor within the exponential term is a measure of diffusion- weighting that is a function of the strength, duration, and temporal spacing of the diffusion-sensitizing gradients.

In equation (2.1), ADC_i is the *ADC* of water along the direction of the diffusion- sensitizing gradient. Higher *ADC* values imply greater rates of water diffusivity along direction *i* in the tissue. *ADC* values can be calculated from DWI by rearranging equation (2.1)

$$ADC_i = -\ln\left(S_i / S_0\right) / b \tag{2.2}$$

If we conceptualize ADC as a tensor in 3-dimensional space, it can be represented by a 3×3 matrix:

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} = R \cdot \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \cdot R^T$$

Where R is the rotation matrix, λ_1, λ_2, and λ_3 are the eigenvalues of diffusion tensor. Figure 2.1 gives an intuition of the information provided by the diffusion tensor

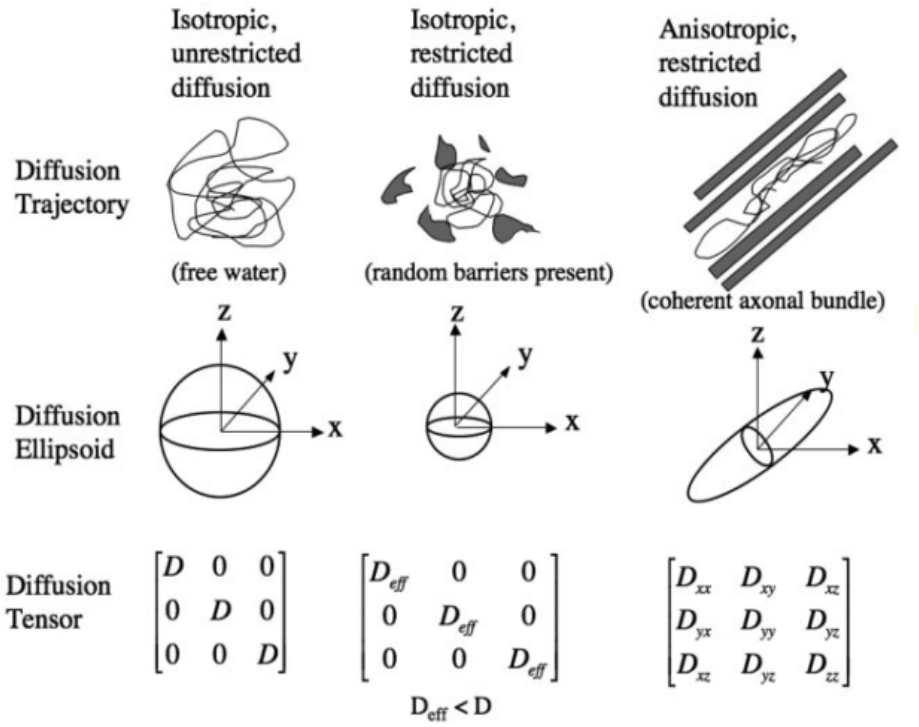


Figure 1 (Mukherjee, 2008). The diffusion ellipsoids and tensors for isotropic unrestricted diffusion, isotropic restricted diffusion, and anisotropic restricted diffusion are shown.

The tensor model was developed to characterize diffusion in anisotropic tissues, where the diffusion coeficient is represented by a 3×3 matrix instead of a scalar due to its directional dependence. Fractional anisotropy (FA) is defined to describe the isotropic character of voxel:

$$\mathrm{FA} = \sqrt{\frac{2}{3}} \cdot \sqrt{\frac{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}{\lambda_1^{\,2} + \lambda_2^{\,2} + \lambda_3^{\,2}}} \qquad (2.3)$$

for eigenvalues λ_1, λ_2, and λ_3 and mean diffusivity $\bar{\lambda}$. FA ranges from 0 (isotropic) to 1 (maximally

anisotropic).

2.2 Probabilistic tractography

The objective of DTI fiber tracking is to determine intervoxel connectivity on the basis of the anisotropic diffusion of water (Mori, et al., 1999). Diffusion weighted imaging is only sensitive to motion of water molecules that is aligned with the motion-probing gradients. In gray matter, since diffusion is roughly isotropic, changing the direction of the gradient does not substantially affect the signal. On the contrary, water diffusion in white matter is often strongly anisotropic and occurs maximally in the same orientation as white matter tracts (Moseley, et al., 1991). The basic idea of tracking fiber tracts with DWI rests on the above fact and the assumption that the dominant direction of axonal tracts can be assumed to be parallel to one of the eigenvectors of the diffusion tensor.

DTI fiber tracking algorithms can be categorized as deterministic and probabilistic.

Fiber assignment by continuous tracking (FACT) is a deterministic method (Basser, et al., 2000). FACT tracks fiber trajectories by following the primary eigenvector from voxel to voxel in 3 dimensional space. Constraints on the maximum curvature of the streamline between successive voxels and on the minimum FA within a voxel for continued propagation of the streamline can be applied in order to guarantee the fiber goes to regions of the brain where the white matter pathways are realistically represented by the diffusion tensor model.

The deterministic tracking technique is limited due to the uncertainty in the fiber orientation caused by noise, patient movement and imaging artifacts (Anderson, 2001).

Probabilistic tractography methods, on the other hand, incorporate the expected uncertainty into the tracking mechanism. Probabilistic tracking techniques tend to disperse trajectories more than deterministic methods and have the potential to delineate a greater portion of a white matter tract.

One of the main confusion of fiber tracking comes from regions of crossing fibers. The typical diffusion tensor model is unable to accurately characterize such regions. Multiple tensor fitting algorithms (Tuch, et al., 2002; Hosey, et al., 2005; Behrens, et al., 2003) provide discrete fiber orientations, but tend to perform poorly when more than two fiber orientations are present. Tournier and his colleagues (2004) introduced a concept of fiber orientation distribution (FOD) or orientation density function (ODF) to model the diffusion attribute of voxels. They estimate FOD using spherical deconvolution (SD) methods. Spherical deconvolution allows more arbitrary FODs than a 3-dimensional tensor model, which enables the probabilistic tractography to be more flexible and reliable

within crossing fiber regions.

With the help of the continuous fiber orientation distribution, the probabilistic tractography algorithm can be described by a model of randomly walking particles (Anwander, et al., 2007). Imagine a particle starting from a seed voxel A, moving in a random manner from voxel to voxel. The transition probability to a neighboring voxel depends on the FOD based on the local diffusivity profile that is modeled from the DTI measurement. This FOD yields higher transitional probabilities along directions with high diffusivity, that is, the presumed fiber directions. Hence, the particle will move with a higher probability along a fiber direction. If we perform this 'trial' many times and count how often particles from voxel A reach a target voxel B, we obtain a relative measure of the probability of tracing a pathway between the 2 voxels.

In this project, we perform probabilistic tractography using the open source package MRtrix (http://www.brain.org.au/software/mrtrix/), which is implemented based on Tournier's approach.

2.3 Gaussian Mixture Model of Tractograms

Researchers tend to propose certain mathematical models to characterize the result of a tractography. Such models of connectivity patterns are usually referred to as "tractograms". In previous research, tractograms were modeled as connectivity histograms or "profiles" (Anwander, et al., 2007) or as "fingerprints" (Beckmann, et al., 2009), defined by the probabilities of connection between a cortical voxel and a set of predefined target regions. Such tractograms can be obtained by running probabilistic tractography repeatedly from seed voxels within ROIs. As a result, the connectivity pattern of a seed voxel is reduced to a vector of probabilities.

The connectivity profile model, however, suffers a limitation that predefined target regions are needed, which requires the prior knowledge of segmentation of brain. This results in a 'chicken-egg' problem in CBP: we need a segmentation of brain in order to obtain the segmentation of brain.

To overcome this drawback, we propose a new representation of tractograms: Gaussian mixture model (GMM). GMMs are parametric probability density functions that have been shown to be useful in approximating densities with complicated shapes (Bishop, 2006). We directly model the tractograms of seed voxels as a distribution of points in 3D image space using GMMs. Figure 2.2 offers an intuitive picture of the distribution of the end of tracks and the reconstructed GMMs of three random seed voxels. The rest of this chapter will give a brief introduction to GMMs and their implementation.

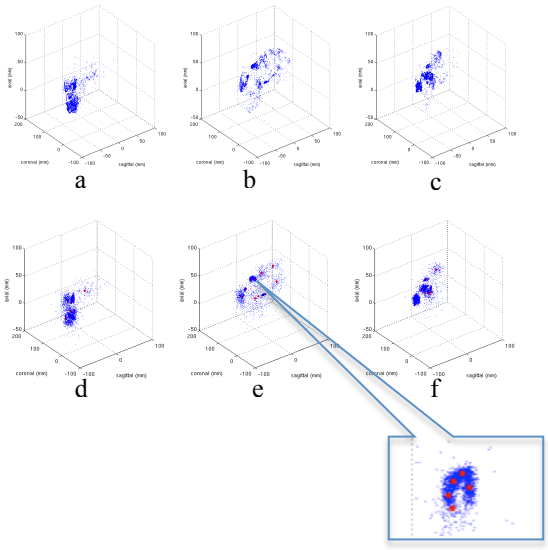


Figure 2.2. Three randomly selected tractograms (top) and their GMMs (bottom). The red dots are the means of the GMM components. 17 components were used to model these sample distributions; parcellation results are relatively insensitive to this number.

2.3.1 Gaussian Mixture Model

The Gaussian mixture model is the most popular mixture model, which is also called mixture of Gaussian (MOG). GMM is a linear superposition of normal distributions. It was developed to provide a richer class of density function than a single Gaussian (Bishop, 2006). In this model, each base distribution of the mixture is a multivariate Gaussian with mean μ_k and covariance Σ_k. Thus the model has the form

$$p(x) = \sum_{k=1}^{K} \pi_k N(x \mid \mu_k, \Sigma_k) \tag{2.4}$$

subject to the constraint

$$\sum_{k=1}^{K} \pi_k = 1, \pi_k \geq 0 \tag{2.5}$$

Each Gaussian in the (2.4) is called a component of the mixture. For the case of a *d* dimensional variable *x,* the Gaussian distribution with its own mean μ_k and covariance Σ_k is defined by

$$N(x \mid \mu_k, \Sigma_k) = \frac{1}{\sqrt{2\pi}} \mid \Sigma_k \mid^{-\frac{d}{2}} \exp(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)) \tag{2.6}$$

The graphical representation of a Gaussian mixture model is shown in Figure 2.3.

The observation $x_i \in X = (x_i, x_2, ..., x_N)$ is statistically independent. The joint conditional density of the data set X is:

$$p(X \mid \Pi, \Theta) = \prod_{i=1}^{N} p(x_i \mid \pi, \vartheta) = \prod_{i=1}^{N} \left[\sum_{k=1}^{K} \pi_k N(x_i \mid \vartheta_k) \right] \tag{2.7}$$

where $\vartheta_k = \{\mu_k, \Sigma_k\}$

Gaussian mixture models provide a practical data-driven, semiparametric framework to approximate unknown distributions. GMM can be viewed as a compromise between (a) single Gaussian for which *K=1,* and (b) the kernel density estimator (Silverman, 1986) for which *K=N,* the number of data points. The single Gaussian distribution is incapable of modeling highly skewed or multimodal data. The kernel density estimator, on the other hand, requires one to retain all *N* data points for the purpose of inference, which is a computationally consuming task. GMMs are popular because theyeffectively parameterize the data, while retaining the accuracy to characterize complex distributions.

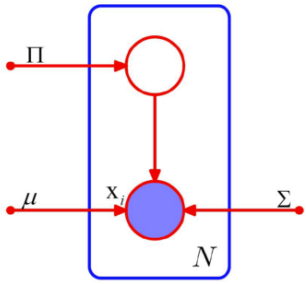


Figure 2.3. The graphical model of a Gaussian mixture model for a set of *N* data points

2.3.2 The Expectation Maximization (EM) Algorithm

In order to find the maximum likelihood for a GMM, a powerful yet elegant tool, the expectation maximization (EM) algorithm, is often used to determine the model parameters (Dempster , et al., 1977; Mclachlan & Krishnan, 2003).

The log-likelihood of GMM, given Eq (2.7) is as follow:

$$\begin{aligned} L(\Pi, \Theta \mid X) &= \ln p(X \mid \Pi, \Theta) \\ &= \ln \prod_{i=1}^{N} p(x_i \mid \pi, \vartheta) = \sum_{i=1}^{N} \ln \left[\sum_{k=1}^{K} \pi_k N(x_i \mid \vartheta_k) \right] \end{aligned} \tag{2.8}$$

Taking the derivatives of $L(\Pi,\Theta \mid X)$ in Eq 2.8, with respect to μ_k and setting to zero, we obtain

$$\frac{\partial L(\Pi,\Theta \mid X)}{\partial \mu_k} = -\sum_{i=1}^{N} \frac{\pi_k N(x_i \mid \mu_k, \Sigma_k)}{\sum_k \pi_k N(x_i \mid \mu_k, \Sigma_k)} \Sigma_k (x_i - \mu_k) = 0 \tag{2.9}$$

Let $z \in \{0,1\}$ be a assignment variable, such that z_k is equal to 1 and all other elements are equal to 0, $\sum_{k=1}^{K} z_k = 1$. We define the joint distribution $p(x,z)$ in terms of a marginal distribution $p(z)$ and a conditional distribution $p(x \mid z)$. The marginal distribution over z is specified in terms of π_k, such that

$$p(z_k = 1) = \pi_k$$

Similarly, the conditional distribution of x given a assignment value for z is a Gaussian

$$p(x \mid z_k = 1) = N(x \mid \mu_k, \Sigma_k) \tag{2.10}$$

And the marginal distribution of x is given by Eq (2.4). Based on Bayes' rule, the posterior probabilities of z given x, denoted by z_{ik}, can be obtained:

$$z_{ik} = p(z_k = 1 \mid x) = \frac{p(z_k = 1) p(x \mid z_k = 1)}{\sum_{k=1}^{K} p(z_k = 1) p(x \mid z_k = 1)} = \frac{\pi_k N(x \mid \mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k N(x \mid \mu_k, \Sigma_k)} \tag{2.11}$$

Note that the posterior probabilities z_{ik} appear in Eq(2.9). Solving Eq(2.9) yields the optimal μ_k

$$\mu_k = \frac{\sum_{i=1}^{N} z_{ik} x_i}{\sum_{i=1}^{N} z_{ik}} \tag{2.12}$$

Once we set the derivative of $L(\Pi,\Theta \mid X)$ with respect to Σ_k, we can solve the optimal Σ_k following the similar manner

$$\Sigma_k = \frac{\sum_{i=1}^{N} z_{ik} (x_i - \mu_k)(x_i - \mu_k)^T}{\sum_{i=1}^{N} z_{ik}} \tag{2.13}$$

Finally, we determine the optimal mixture fraction π_k by maximizing $L(\Pi,\Theta \mid X)$. Considering the constraint of mixture faction in Eq(2.5), maximize $L(\Pi,\Theta \mid X)$ plus a Lagrange multiplier:

$$L(\Pi,\Theta \mid X) + \lambda(\sum_{k=1}^{K} \pi_k - 1) \tag{2.14}$$

Maximizing Eq (2.14) gives

$$\lambda = -\sum_{i=1}^{N} \frac{N(x_i \mid \mu_k, \Sigma_k)}{\sum_k N(x_i \mid \mu_k, \Sigma_k)} = -\sum_{i=1}^{N} \frac{z_{ik}}{\pi_k} \tag{2.15}$$

Replacing this result with λ in Eq(2.14) we obtain

$$\pi_k = \frac{1}{N} \sum_{i=1}^{N} \frac{N(x_i \mid \mu_k, \Sigma_k)}{\sum_k N(x_i \mid \mu_k, \Sigma_k)} \tag{2.16}$$

Note the maximum likelihood results (2.12), (2.13), and (2.16) are not close-form solutions to the parameters of the mixture model. However, the EM algorithm describes an iterative procedure for estimating the parameters, which maximize the likelihood of the currently available data. The iterative scheme works as follows: first we choose some initial values of the parameters; then we alternatively update them with the E step and M step. In the E step, namely expectation step, we evaluate the posterior probabilities with current estimate of parameters, using Eq (2.11). Then for the M step, or maximization step, we re-estimate and update of the mean, covariance, and mixture fraction using (2.12), (2.13), and (2.16). McLachlan and Krishnan (2003) have shown that each of the E and M step is guaranteed to increase the log-likelihood function, $L(\Pi, \Theta \mid X)$. And the algorithm is guaranteed to converge to a local optimum of the mixture model. In practice, we repeat the alternate updates until the difference between two consecutive estimates falls below certain threshold.

In general cases the E step and M step can be identified as:

- E step: Given the present parameter estimate $\Theta^{(t)}$, the expectation of the posterior distribution is

$$U(\theta; \theta^{(t)}) = E[\sum_Z p(Z \mid X, \theta^{(t)}) \ln p(Z \mid X, \theta)] \tag{2.17}$$

- M step: Update the estimate for the parameters $\Theta^{(t+1)}$, by maximizing the expectation form

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} U(\theta; \theta^{(t)}) \tag{2.18}$$

With the help of EM algorithm, parameters of the Gaussian mixture model can be quickly inferred. We, however, also notice that the estimate of EM algorithm is only guaranteed to converge to the local stationary point, instead of a global optimum. Thus, for a likelihood function with multiple stationary points, either local or global possibly, the convergence depends on the initial choice of the parameters. In this project, using the Gaussian mixture pipeline provided by MATLAB©, we apply two means to increase the chance for reaching the global optimum: first we select the initial parameters randomly from the data set, and spread them as possible; second we run the algorithm for several times and

choose the parameters resulting in the maximum likelihood. Once we obtain the GMM presentation of the tractograms, we calculate the pairwise distance between the GMMs of voxels. The distance metrics we use is the earth mover's distance (EMD) (Rubner, et al., 2000). The EMD corresponds to the cost of moving 'material' from one distribution to another, and is computed using linear programming methods for transportation problems. The distributions can be sampled from the GMM of the tractogram rather than the tractogram itself. Applying the distance metrics, we thus obtain a distance matrix, which can be used for clustering.

Chapter 3 CLUSTERING TECHNIQUES

Three clustering techniques will be introduced in this chapter: spectral clustering, hidden Markov random fields model, and non-parametric models. The three methods apply to particular clustering cases based on their characteristics. And in some situation more than one method are used simultaneously.

3.1 Spectral Clustering

The nature of connectivity-based parcellation can be viewed as the clustering of voxels of ROI in a high-dimensional 'connectivity space'; tractograms, or connectivity patterns, are the features in this space. It is hard to identify these 'points' with a vector as we do in a Euclidean space, but we are able to define and compute the pairwise distance (similarity) between points as explained in the previous chapter. Spectral clustering methods have the benefit of allowing us to include arbitrary features for representing data. It is classified as a distance-based clustering approach. They assume that the data lie on a low-dimensional manifold but are represented in a high-dimensional feature space (Socher, et al., 2011). As a result of this property, spectral clustering and its derivatives have become one of the emerging techniques for medical image segmentation. In order to recover the underlying cluster structure they perform the following steps. First, the connectivity features of each voxel are used to compute a pairwise distance (similarity) matrix. Second, this matrix is then used to map the observations from this implicit representation into a lower dimensional Euclidean space. Third, in this space most methods apply model-based clustering methods such as k-means, fixing the number of clusters by hand (Luxburg, 2007). In this project, after conducting the first two steps above, the segmentation is completed with various model-based clustering methods according to the purpose of the experiment. The following description of spectral clustering is mainly borrowed from Luxburg's tutorial in 2007.

3.1.1 ***Similarity Graphs***

Given a set of data points $x_1, \ldots, x_n$ and some notion of similarity $s_{ij} \geq 0$ between all pairs of data points

x_i and x_j, the data set can be represented in a graphical point of view. We define a *similarity graph* $G=(V, E)$. Each vertex v_i represents a data point x_i; the edge is weighted by s_{ij}, if s_{ij} is strictly positive. Thus the problem of clustering can now be reformulated using the similarity graph: we want to find a partition of the graph such that the edges between different groups have very low weights (which means that points in different clusters are dissimilar from each other) and the edges within a group have high weights (which means that points within the same cluster are similar to each other). The partition of the graph is often called a *cut*. Figure 3.1 gives an illustration of this process.

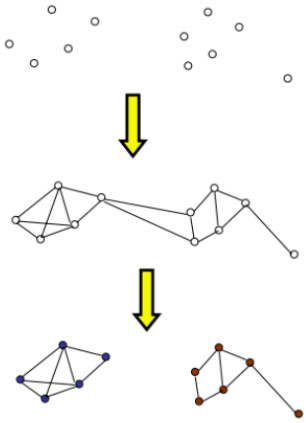


Figure 3.1. First arrow: graph representation of the data point. Second arrow: partitioning the graph by finding the minimum cuts (Image from Liu's lecture)

Let $G = (V, E)$ be an undirected weighted graph with vertex set $V = \{v_1, \ldots, v_n\}$, the weight is denoted by $w_{ij} \geq 0$. The weighted adjacency matrix of the graph is the matrix $W = (w_{ij})_{i,j=1,\ldots,n}$. If $w_{ij} = 0$ this means that the vertices v_i and v_j are not connected by an edge. As G is undirected we require $w_{ij} = w_{ji}$, which means the adjacency matrix is symmetric. The degree of a vertex $v_i \in V$ is defined as

$$d_i = \sum_{j=1}^{n} w_{ij} \tag{3.1.1}$$

The degree matrix D is defined as the diagonal matrix with the degrees $d_1, \ldots, d_n$ on diagonal. For two not necessarily disjoint sets $A, B \subset V$ we define

$$W(A,B) = \sum_{i \in A, j \in B} w_{ij} \tag{3.1.2}$$

We consider two different ways of measuring the size of a subset $A \subset V$

$|A|$: = the number of vertices in A

$$vol(A) := \sum_{i \in A} d_i = \sum_{i \in A} \sum_{j=1}^{n} w_{ij}$$

$|A|$ is called cardinality of A, while *vol(A),* or volume of A, measures the size of A by summing over the weights of all edges attached to vertices in A.

3.1.2 Graph Laplacian

Spectral clustering is motivated by spectral graph theory (Chung, 1997), whose major tool is the graph Laplacian matrix.

Unnormalized: The unnormalized graph Laplacian matrix of $G = (V, E)$ is defined as

$$L = D - W \tag{3.1.3}$$

The matrix L has the following properties:

1. L is symmetric and positive semi-definite.

2. L has n non-negative, real-valued eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, λ_1*'s* corresponding eigenvector is the constant one vector ***1.***

3. For every vector $f \in \Re^n$ we have

$$f'Lf = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2 \tag{3.1.4}$$

4. Invariance to self-edges. The unnormalized graph Laplacian does not depend on the diagonal elements of the adjacency matrix W, for $L_{ij} = \begin{cases} d_i - w_{ii}, (i = j) \\ -w_{ij}, (i \neq j) \end{cases}$

The unnormalized graph Laplacian and its eigenvalues and eigenvectors can be used to describe many properties of graphs (Mohar, 1997). One significant example is the following:

Proposition 1 (Number of connected components and the spectrum of L) Let G be an undirected graph with non-negative weights. Then the multiplicity k of the eigenvalue 0 of L equals the number of connected components $A_1, \ldots, A_k$ in the graph. The eigenspace of eigenvalue 0 is spanned by the indicator vectors $\mathbf{1}_{A1}, \ldots, \mathbf{1}_{Ak}$ of those components.

This proposition is extremely useful in practice. For example, one can decide the number of clusters by finding the eigenvalues of L that are equal or very close to zero.

Normalized: There are two matrices which are called normalized graph Laplacians in the literature. Both matrices are closely related to each other and are defined as

$$L_{sym} := D^{-1/2}LD^{-1/2} = I - D^{-1/2}WD^{-1/2} \tag{3.1.5}$$

$$L_{rw} := D^{-1}L = I - D^{-1}W \tag{3.1.6}$$

We denote the first matrix by L_{sym} as it is a symmetric matrix, and the second one by L_{rw} as it is closely related to a random walk. In the following we summarize several properties of L_{sym} and L_{rw}.

Proposition 2 (Properties of L_{sym} and L_{rw}) The normalized Laplacians satisfy the following properties:

1. For every $f \in \Re^n$ we have

$$f' L_{sym} f = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}\left(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}}\right)^2 \tag{3.17}$$

2. λ is an eigenvalue of L_{rw} with eigenvector u if and only if λ is an eigenvalue of L_{sym} with eigenvector $w = D^{1/2} u$.

3. λ is an eigenvalue of L_{rw} with eigenvector u if and only if λ and u solve the generalized eigen problem $Lu = \lambda Du$.

4. 0 is an eigenvalue of L_{rw} with the constant one vector $\mathbf{1}$ as eigenvector. 0 is an eigenvalue of L_{sym} with eigenvector $D^{1/2}\mathbf{1}$.

5. L_{sym} and L_{rw} are positive semi-definite and have n non-negative real-valued eigenvalues $0 = \lambda_1 \leq \ldots \leq \lambda_n$.

As it is the case for the unnormalized graph Laplacian, the multiplicity of the eigenvalue 0 of the normalized graph Laplacian is related to the number of connected components:

Proposition 3 (Number of connected components and spectra of L_{sym} and L_{rw}) Let G be an undirected graph with non-negative weights. Then the multiplicity k of the eigenvalue 0 of both L_{rw} and L_{sym} equals the number of connected components $A_1, \ldots, A_k$ in the graph. For L_{rw}, the eigenspace of 0 is spanned by the indicator vectors $\mathbf{1}_{Ai}$ spanned by the vectors $D^{1/2}\mathbf{1}_{Ai}$.

3.1.3 Spectral Clustering Algorithm

Suppose we would like to cluster n data points $x_1, \ldots, x_n$, and a similarity matrix $S = \{s_{ij}\}$, $s_{ij} \geq 0$. We describe the spectral clustering algorithm based on the settings above.

Input: Similarity matrix $S \in \Re^{n \times n}$, number of clusters k.
• Construct a similarity graph by one of the ways described in Section 2. Let W be its weighted adjacency matrix.
• Compute the Laplacian: unnormalized L (3.13) or normalized L_{sym} (3.15) or L_{rw} (3.16).
• Compute the first k eigenvectors $u_1, \ldots, u_k$ of L (L_{sym}, L_{rw}).
• Let $U \in \Re^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.
• For $i = 1, \ldots, n$, let $y_i \in \Re^k$ be the vector corresponding to the i-th row of U.
• Cluster the points $(y_i)_{i=1,\ldots,n}$ with the k-means algorithm into k clusters $C_1, \ldots, C_k$.
Output: Clusters $C_1, \ldots, C_k$

A pair-wise distance matrix of the connectivity patterns and the matrix after applying spectral clustering are shown in Figure 3.2. The ROI is right hemisphere inferior parietal lobule (IPL) of one subject with 797 voxels. The Laplacian used the symmetric normalized Laplacian.

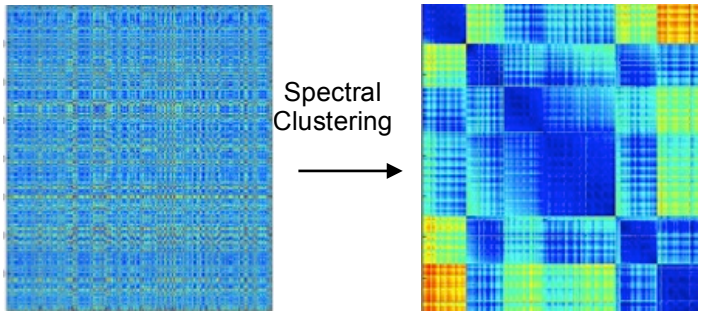
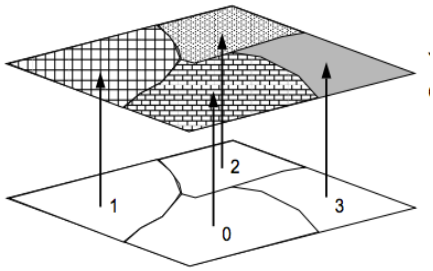


Figure. 3.2. Dissimilarity matrices of 797 voxels from 1 subject. (a) Matrix of the pair-wise EMD distance between connectivity patterns of seed voxels; (b) Reordered matrix after performing spectral clustering on the voxels.

3.2 Hidden Markov Random Fields Model

The Markov random field (MRF) model has gained huge popularity in computer vision problems, such as denoising (Besag, 1986), restoration (Geman & Geman, 1984), and segmentation (Li, 1995). Since Zhang, et al. (2001) introduced a derivative of MRF, hidden Markov random field (HMRF) model into brain MR image segmentation, HMRF has proved itself particularly efficient for the problems in medical imaging. A Hidden Markov random field model is a graphical probability model in which the true states are not observed but are estimated indirectly through an observation field. In an HMRF model, we assume that the spatial information is encoded via the interaction of adjacent sites; the segmentation of the image, although unobservable, can be indirectly estimated from the observations

(Zhang, 2001). Figure 3.3 illustrates the implication of HMRF.



Y: features observed from the image

X: underlying field containing the labeling information of each

Figure 3.3. Intuition of HMRF. The values of voxels in the image space (*Y)* are encoded through the mutual influences of neighboring voxels. The classification of voxels (*X),* which is latent, can be inferred from the observable features

3.2.1 MRF Theory

The most used imaging encoded model is the *contextual constraint*. MRF theory provides a convenient and consistent way to model context-dependent entities such as image pixels and correlated features (Zhang, 2001).

In an MRF, the sites ***S*** are related to one another via a neighborhood system $N = \{N_i, i \in S\}$, where N_i denotes the sites adjacent to voxel *i*. A stochastic field *X* is said to be an Markov random field if for all $i \in S$

$$P(x_i \mid x_{S-\{i\}}) = P(x_i \mid x_{N_i}) \tag{3.2.1}$$

Eq (3.2.1) implies that in a MRF, the value of voxel $i \in S$ is independent from the neighboring voxels given its neighborhoods. MRF theory generates from the theorem of Markov chain. This thought significantly reduces the dependencies within a random field by taking advantage of the conditional independency, which in practice leads to tremendous decrease in computational complexity.

According to the Hammersley–Clifford theorem (Besag, 1974), an MRF can equivalently be characterized by a Gibbs distribution:

$$P(x) = Z^{-1} \exp(-U(x)) \tag{3.2.2}$$

where *Z* is a normalization constant and *U(x)* is an potential function.

3.2.2 HMRF Model

The concept of hidden Markov random field model is derived from Markov random field (MRF) model and hidden Markov models (HMM), which are defined as stochastic processes generated by a Markov chain whose state sequence cannot be observed directly, only through a sequence of observations. Each observation is assumed to be a stochastic function of the state sequence. The underlying Markov chain changes its state according to a transition function. HMRF extends HMM from a 1-dimensional Markov chain to higher-dimensional spaces; extends MRF by introducing hidden, or latent, variables.

In a HMRF model, there's an *observable random field* $y = (y_1, \ldots, y_N)$ where each y_i is the feature value of a voxel in our problem. The goal is to infer a *hidden random field* $x = (x_1, \ldots, x_N)$ where $x_i \in L$. In the image segmentation context, x_i *is* a configuration of labels and L is the set of all possible labels (e.g. cortical fields in our application).

Conditional independence: Besides the conditional independencies in (3.2.1), an HMRF also holds conditional independencies between x and y:

$$P(y \mid x) = \prod_{i \in S} P(y_i \mid x_i) \tag{3.2.3}$$

Thus the joint probability is

$$P(y,x) = P(y \mid x)P(x) = P(x)\prod_{i \in S} P(y_i \mid x_i) \tag{3.2.4}$$

Given x_i*'s* neighborhood configuration x_{Ni}, according to the local conditional independencies of MRFs, (3.2.4) can be rewritten as

$$P(y_i, x_i \mid x_{N_i}) = P(y_i \mid x_i)P(x_i \mid x_{N_i}) \tag{3.2.5}$$

This way the joint distribution of (y, x) has been highly decoupled. If we further the assume that y is drawn from a probability function $f(y;l,\theta)$, with parameter set θ and $l \in L$, we can obtain the marginal distribution of y_i

$$p(y_i \mid x_{N_i}, \theta) = \sum_{l \in L} p(y_i, l \mid x_{N_i}, \theta) = \sum_{l \in L} f(y_i;\theta_l)p(l \mid x_{N_i}) \tag{3.2.6}$$

3.2.3 Estimating HMRF Using EM Algorithm

According to the MAP criterion, we seek the labeling x^* which satisfies

$$x^* = \arg\max_x \{P(y \mid x, \Theta) P(x)\} \tag{3.2.7}$$

The prior probability *P(x)* is a Gibbs distribution (3.2.2), and the joint likelihood probability is

$$P(y \mid x, \Theta) = \prod_i P(y_i \mid x, \Theta) = \prod_i P(y_i \mid x_i, \theta_{x_i}) \tag{3.2.8}$$

where $P(y_i \mid x_i, \theta_{x_i})$ is a Gaussian distribution with parameters $\theta_{x_i} = (\mu_{x_i}, \Sigma_{x_i})$.

Similar to the inference of the Gaussian mixture models, one can also estimate the MAP parameters of HMRF using an EM algorithm. Following the procedures of EM described in Chapter 2, we iteratively update the estimates as below:

1. Start with initial parameter set $\Theta^{(0)}$.
2. Calculate the likelihood distribution $P^{(t)}(y_i \mid x_i, \theta_{x_i})$ given by (3.2.8)
3. Using the current parameter set $\Theta^{(t)}$ to estimate the labels by MAP estimation:

$$\begin{aligned} x^{(t)} &= \operatorname*{argmax}_{x \in X} \{P(y \mid x, \Theta^{(t)}) P(x)\} \\ &= \operatorname*{argmin}_{x \in X} \{U(y \mid x, \Theta^{(t)}) + U(x)\} \end{aligned} \tag{3.2.9}$$

   The implementation of the MAP estimation is discussed in the next section

4. Compute the posterior distribution for all possible $l \in L$ and all voxels y_i:

$$P^{(t)}(l \mid y_i) = \frac{N(y_i; \theta_l) P(l \mid x_{N_i}^{(t)})}{P^{(t)}(y_i)} \tag{3.2.10}$$

   where $N(y_i; \theta_l)$ is the Gaussian distribution of y_i with parameter $\theta_l = (\mu_l, \Sigma_l)$.

   The marginal probability

$$P^{(t)}(y_i) = \sum_{l \in L} N(y_i; \theta_l) P(l \mid x_{N_i}^{(t)}) \tag{3.2.11}$$

   Here we have

$$P(l \mid x_{N_i}^{(t)}) = \frac{1}{Z} \exp(-\sum_{j \in N_i} V_c(l, x_j^{(t)})) \tag{3.2.12}$$

   $V_c(l,x)$ is called the *clique potential* of field *x*, which is discussed in the next section

5. Use $P(l \mid x_{N_i}^{(t)})$ to update the parameters of Gaussian as in (2.12) and (2.13)

3.2.4 MAP Estimation

In (3.2.9) we introduce the total posterior energy in order to find *x's* MAP estimate:

$$x = \underset{x \in X}{\operatorname{argmin}} \{U(y \mid x, \Theta) + U(x)\} \tag{3.2.13}$$

The total posterior energy has two compositions: the likelihood energy and the prior energy. the likelihood energy is defined by

$$U(y \mid x, \Theta) = \sum_i U(y_i \mid x_i, \Theta) = \sum_i [\frac{1}{2}(y_i - \mu_{x_i})\Sigma_{x_i}^{-1}(y_i - \mu_{x_i})^T + \ln \Sigma_{x_i}] \tag{3.2.14}$$

It's easy to see that the likelihood energy is proportional to the minus log-likelihood of the *P(x|y):*

$\log P(x \mid y) \propto -U(x \mid y)$

Note that $U(x \mid y) = U(y \mid x) + U(x) + c$ is the posterior energy, where *c* is a constant and the prior energy function *U(x)* has the form

$$U(x) = \sum_{c \in C} V_c(x) \tag{3.2.15}$$

which is the sum of clique potentials $V_c(x)$ over all the cliques *C*. A clique is defined as a subset of sites in in which every pair of distinct sites are neighbors. The value of clique potential depends on the local configuration and physical properties of the clique. In the context of image segmentation, a clique potential is usually used to characterize the influence from adjacent voxels on voxel *i*. For this project, we define the clique potential of voxel *i* as the number of its adjacent voxels who have different labels:

$$V_c(x_i) = \frac{1}{2}(1 - I_{x_i, x_j}). \tag{3.2.16}$$

where $x_j \in x_{N_i}$, and

$$I_{x_i, x_j} = \begin{cases} 0, & x_i \neq x_j \\ 1, & x_i = x_j \end{cases} \tag{3.2.17}$$

3.3 Non-parametric Bayes Models

So far we have discussed a distance-based, or cost-based clustering technique: spectral clustering; and a model-based clustering approach: hidden Markov random field model. The principal problem with these two methods s how to decide the number of clusters *K*. A recently very active field study provides some ways to tackle this problem, that is the *non-parametric Bayes model*. The non-parametric Bayes models also feature in mixture models, but unlike the Gaussian mixture models, they are *infinite mixture models,* which allows the number of clusters to grow as the population of data increases. They don't impose a priori knowledge of the cluster number *K*.

In this section, I will give a brief introduction to a typical non-parametric Bayes model, the Dirichlet

process model (DPM).

3.3.1 The Dirichlet Process

The Dirichlet process (DP) was first developed in Ferguson (1973). The Dirichlet process is a distribution over distributions. It is parameterized by a concentration parameter $\alpha > 0$ and a base distribution H, which is a distribution over a space Θ. A random variable drawn from a DP is itself a distribution over Θ. A random distribution G drawn from a DP is denoted $G \sim \mathrm{DP}(\alpha, H)$.

The DP is defined implicitly by the requirement that $(G(T_1),\ldots,G(T_K))$ has a joint Dirichlet distribution:

$$(G(T_1),\ldots,G(T_K)) \sim \mathrm{Dir}(\alpha H(T_1), \ldots, \alpha H(T_K)) \qquad (3.3.1)$$

This means that if draw a random distribution from the DP and add up the probability mass in a region $T \in \Theta$, then there will on average be $H(T)$ mass in that region. The concentration parameter plays the role of an inverse variance; for higher values of α, the random probability mass $G(T)$ will get more concentrated around $H(T)$.

A DP can be viewed from two perspectives: an infinite mixture model and distribution drawn from a Dirichlet process. The graphical models in Figure 34 illustrate these two views:

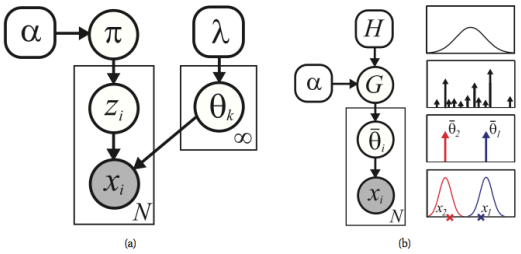


Figure 3.4 (Sudderth, 2006) Two views of a DP mixture model. Left: infinite number of clusters parameters, θ_k, and $\pi \sim \mathrm{GEM}(\alpha)$. Right: G is drawn from a DP.

3.3.2 Chinese Restaurant Process

The non-parametric Bayes approach addresses the problem of deciding the number of clusters by assuming that it is infinite, while specifying the prior over infinite groupings in such a way that it favors assigning data to a small number of groups. The prior over groupings is called the Chinese restaurant process (CRP; Pitman, 2002), a distribution over infinite partitions of the integers. The

analog is as follows. Imagine a Chinese restaurant with an infinite number of tables, and imagine a sequence of customers entering the restaurant and sitting down. The tables are like clusters, and the customers are like observations. The first customer enters and sits at the first table. The second customer enters and sits with the first person with probability $\frac{1}{1+\alpha}$, or at another table, with probability $\frac{\alpha}{1+\alpha}$. When the *N*th customer enters the restaurant, he may join an occupied table *k* with probability proportional to the number of people already sitting there (denoted N_k), otherwise, he may choose to sit at another table *k** with probability proportional to α. Figure 3.5 shows a scheme of this process.

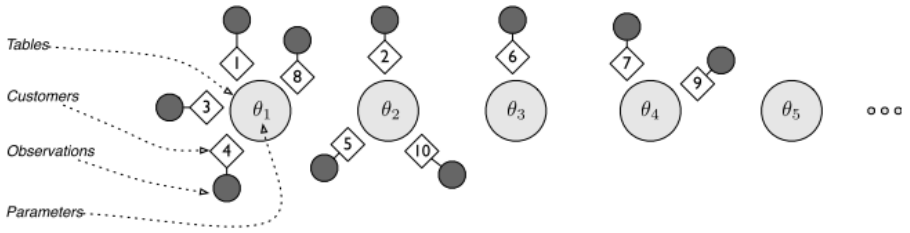


Figure. 3.5 (Gershman & Blei, 2012). The Chinese restaurant process. The generative process of the CRP, where numbered diamonds represent customers, attached to their corresponding observations (shaded circles). The large circles represent tables (clusters) in the CRP and their associated parameters (θ). Technically the parameter values {θ} are not part of the CRP per se, but rather belong to the full mixture

To formalize process, we let z_N be the table assigned to the *Nth* coming customer. A draw from this distribution can be generated by sequentially assigning data points to classes with probability below:

$$P(z_N = k \mid z_{1:N-1}) = \begin{cases} \frac{N_k}{N-1+\alpha} & if k \le K \\ \frac{\alpha}{N-1+\alpha} & otherwise \end{cases} \tag{3.3.2}$$

α in (3.3.2) is the concentration parameter. Obviously, a larger α will produce higher number of clusters.

CRP results in a fact that large clusters tend to get even larger. This is a power rule often referred to as Matthew Effect. CRP also demonstrates an important property: the assignment of clusters is *exchangeable*. This means that the order of the customers entering the restaurant won't affect the table assignment scheme at last. The implication of exchangeability is very deep. This thesis doesn't cover detailed discussion about it. The CRP constructs an important sampling scheme of DPM. There are

other constructions such as *stick breaking construction* and *Polya urn process*. DPM is implemented based on CRP in the scope of this thesis.

3.3.3 Fitting The Dirichlet Process Model

The most widely used posterior inference methods in nonparametric Bayes models are Markov Chain Monte Carlo (MCMC) methods. The idea of MCMC methods is to define a Markov chain on the hidden variables that has the posterior as its equilibrium distribution (Andrieu, et al., 2003). By drawing samples from this Markov chain, one eventually obtains samples from the posterior. A simple form of MCMC sampling is Gibbs sampling, where the Markov chain is constructed by considering the conditional distribution of each hidden variable given the others and the observations (Gershman, 2012). Thanks to the exchangeability property described in Section 3.3.2, CRP mixtures are particularly amenable to Gibbs sampling—in considering the conditional distributions, each observation can be considered to be the ‘‘last’’ one and the distribution of (3.3.2) can be used as one term of the conditional distribution. The pseudocode of the Gibbs sampling implementation in Figure 3.6 is borrowed from Neal (2000). This algorithm tend to perform well and efficiently. DPM avoids getting stuck in poor local optima as other model-based algorithms, because DPM creates redundant clusters at the first stage, which enables it to escape local stationary points.

Although DPM demonstrates its capability in adaptively clustering data points, it’s still an issue that the tuning hyper-parameters (α, for instance) can be tricky. Sometimes it could be as hard as finding the right number of clusters. Experiments were conducted to investigate the impact of hyper-parameters on the connectivity-based parcellation with our image data. The results are discussed in chapter 6.

for *each* $i = 1 : N$ *in random order* **do**
  Remove $\mathbf{x}_i$'s sufficient statistics from old cluster z_i ;
  for *each* $k = 1 : K$ **do**
    Compute $p_k(\mathbf{x}_i) = p(\mathbf{x}_i|\mathbf{x}_{-i}(k))$;
    Set $N_{k,-i} = \dim(\mathbf{x}_{-i}(k))$;
    Compute $p(z_i = k|\mathbf{z}_{-i}, \mathcal{D}) = \frac{N_{k,-i}}{\alpha+N-1}$;
  Compute $p_*(\mathbf{x}_i) = p(\mathbf{x}_i|\boldsymbol{\lambda})$;
  Compute $p(z_i = *|\mathbf{z}_{-i}, \mathcal{D}) = \frac{\alpha}{\alpha+N-1}$;
  Normalize $p(z_i|\cdot)$;
  Sample $z_i \sim p(z_i|\cdot)$;
  Add $\mathbf{x}_i$'s sufficient statistics to new cluster z_i ;
  If any cluster is empty, remove it and decrease K;

Figure 3.6. Collapsed Gibbs sampler for DPM (Neal, 2000)

Chapter 4 COMPRESSED SENSING CROSSING FIBERS RESOLUTION

In order to characterize the connectivity of human brain in vivo, the usual diffusion tensor model, is inadequate in crossing fiber regions since it assumes a single fiber orientation for each voxel. Several high angular resolution diffusion-weighted imaging (HARDI) based techniques, such as spherical deconvolution (SD, Tournier, et al., 2004) and q-ball imaging (QBI, Tuch, et al., 2004), have been proposed to address this issue, but they are problematic for either time consumption or hardware constraints. Recently a new compressed sensing (CS, Landman, et al., 2011) approach has been presented. In this chapter, we introduce the basic of compressed sensing and Landman's method.

4.1 Compressed Sensing

In recent years, compressed sensing (CS) has attracted considerable attention in areas of mathematics and engineering. CS builds upon the fundamental fact that we can represent many signals using only a few non-zero coefficients in a suitable basis or dictionary. Nonlinear optimization can then enable recovery of such signals from very few measurements. One of the most popular techniques for signal compression is known as transform coding, and typically relies on finding a basis or frame that provides sparse or compressible representations for signals in a class of interest, *i.e., transform sparsity* (Bruckstein, et al., 2009). By a sparse representation, we mean that for a signal of length N, we can represent it with $K << N$ nonzero coefficients. Candès, Romberg and Tao (2006) and Donoho (2006) have elaborated the compressed sensing (or compressive sampling) approach, whose essential idea is: instead of compressing after sampling in a high rate, one may want to directly sense the signals in a compressed form. Candès, et al. and Donoho showed that a finite-dimensional signal can be recovered from a small set of linear measurements, as long as it satisfies three requirements: (a) the desired signal satisfies transform sparsity, (b) the transform representation of data is incoherent (noise like), (c) a non-linear reconstruction is used to enforce both sparsity of the image representation and consistency with the acquired data (Lustig, 2008).

Consider a real-valued, N-length, one-dimensional signal x, which can be viewed as an $N \times 1$ column vector in $\Re^N$ with sparsity K. Consider a general linear measurement process that computes $M < N$

inner products between x and a (fixed) collection of vectors $\{\phi_j\}_{j=1}^{M}$ as in $y_j = \langle x, \phi_j \rangle$. Arrange the measurements y_j in an $M \times 1$ vector y and the measurement vectors ϕ_j^T in an $M \times N$ matrix Φ. Then, y can be written as

$$y = \Phi x \tag{4.1.1}$$

To recover x from a fewer measurements y, the problem consists of designing a) a stable measurement matrix Φ such that the salient information in any K-sparse or compressible signal is not damaged by the dimensionality reduction and b) a reconstruction algorithm.

Candès et al.,(2006) and Donoho (2006) have demonstrated that for the matrix Φ which is incoherent with high probability, more specifically, an $M \times N$ *iid* Gaussian matrix, x can be recovered with high probability if $M \geq cK \log(N/K)$, with c a small constant. There for the designing of the transform matrix at least requires that the number of the random measurements $M \geq cK \log(N/K) << N$.

The general idea of reconstruction algorithm is to penalize non-zero x using minimum norm constraints.

- **Minimum l_2 norm reconstruction**

  The classical approach to regularization problems is to find the vector in the translated null space with the smallest l_2 norm (energy) by solving

  $$\hat{x} = \operatorname{argmin} \|\Phi x - y\|_2 + \beta \|x\|_2 \tag{4.1.2}$$

  This optimization is convenient because it has analytical. Unfortunately, however, the l_2 minimization will not find a K-sparse solution.

- **Minimum l_0 norm reconstruction**

  The l_0 norm counts the number of non-zero entries in signal, which exactly accords the motivation of the regularization. The optimization question thus becomes

  $$\hat{x} = \operatorname{argmin} \|\Phi x - y\|_2 + \beta \|x\|_0 \tag{4.1.3}$$

  Unfortunately, it is still problematic because the solution to the l_0 norm is both numerically unstable and NP-complete (Baraniuk, 2007).

- **Minimum l_1 norm reconstruction**

  One of the surprisingly results deriving from CS is that the l_1 norm

  $$\hat{x} = \operatorname{argmin} \|\Phi x - y\|_2 + \beta \|x\|_1 \tag{4.1.4}$$

  provides almost as desirable results as the l_0 norm does. This is a numerically tractable convex optimization problem that can be solved by conjugate gradient approaches (Kim, 2007) or basis pursuit methods (Candès, 2006)

4.2 Resolving Crossing Fibers Using CS

In 2008, Landman et al. proposed a crossing fibers resolution method using compressed sensing, which they called Crossing Fiber Angular Resolution of Intra-Voxel structure (CFARI). The basic idea of this approach is to characterize a DW signal as a sparse linear combination of a set of selected basis of possible fiber directions, then reconstruct the real fiber orientations from the basis.

CFARI models each voxel as a discontinuous collection of compartments where each compartment contains one or more bundles of neuronal fibers, each bundle being modeled by a traditional diffusion tensor model as we discussed in Chapter 2. The objective is to determine the fractional contributions and directions for each compartment by fitting the model to the measured data. For the DTI measures, the observed signal intensity S_k at a voxel is a mixture of signals, each one of which is described by the Stejskal - Tanner tensor model

$$\frac{S_k}{S_0} = \sum_i^N f_i e^{-b g_k^T D_i g_k} + \eta \tag{4.2.1}$$

where S_0 is a the intensity on the $b=0$ image (no diffusion weighting), g_k is the diffusion gradient, b is the diffusion sensitization strength, N is the number of compartments within a voxel, f_i is the mixture fraction of each compartment (unknown), D_i is the tensor matrix associated with the i^{th} compartment, and η is the noise term.

If we have a *reconstruction basis* of tensors at orientations which are uniformly distributed over a sphere, we can regard the net response of a voxel with crossing fibers as a linear combination of the basis vector set. Further we expect the required number of compartments required to model the response will be small. We apply compressed sensing techniques to reconstruct the mixture fraction f_i of the basis from the DTI signals with a sparse representation. We define $y_k=S_k/S_0$, thus Eq. (4.2.1) has a matrix form:

$$y = Sf + \eta \tag{4.2.2}$$

where S is the $K\times N$ matrix comprising the attenuation terms of the reconstruction basis evaluated at the K measurement directions.

Similar to Eq (4.1.4), we can write the compressed sensing criterion of Eq. (2) for the estimate of f :

$$f = \text{argmin}_{f\geq 0} \left\| Sf - y \right\|_2^2 + \beta \left\| f \right\|_1 \tag{4.2.3}$$

In (4.2.3), β is a sparsity regularization parameter controlling the trade-off between the precision of model fitting (the l_2 norm) and the sparsity requirement (l_1 norm). As β approaches zero, the

optimization problem reduces to ordinary least-squares regression. As β grows, the sparsity term dominates, which may lead to very small f values. The choice of β can significantly influence the results estimates.

Our implementation solves the convex optimization problem in (4.2.3) with an inferior-point method (Kim, 2007) that uses a preconditioned conjugate gradient method to compute the search step.

Chapter 5 METHODOLOGY

This chapter details the pipeline we adopted for the single and multi – subject CBP and the compressed sensing based resolution of crossing fiber regions

5.1. Material And Pre-processing

The data were obtained using single-shot DWI with 64 unique gradient directions, b-value of 1000mm/s^2 and 2 mm isotropic resolution. 16 healthy, adult subjects (at the age ranging from 21 to 53, 8 males and 8 females) were scanned with a Siemens Trio 3T scanner. Noise and eddy current distortions were removed from DW images using FSL (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/; Smith, et al. 2004). Then a 2 mm thick cortico-subcortical surface was extracted in each subject's T1-weighted image with FreeSurfer (http://surfer.nmr.mgh.harvard.edu/fswiki, Dale, et al.,1999). We mapped these segmentations into the subject's DWI space and used the voxels along that surface ribbon which lay at the gray-white matter interface to define both the target regions and seed voxels for probabilistic tractography.

5.2 Generating Tractography

Probabilistic tractography was performed using constrained spherical deconvolution with MRtrix (Tournier, et al., 2007). Tractograms were formed by generating 5000 tracks from each seed voxel and retaining only those tracks that reached the gray-white matter interface. A minimum length of 20 mm is set to be the constraint when generating the tracks to exclude pathways that end too quickly. The distributions of end points of the retained tracks form the tractograms, one per seed voxel. For multi-subject analysis, brains were mapped to 3D MNI (Beckmann, et al., 2009) space using FSL.

5.3 Fitting the Gaussian Mixture Model

We directly model the tractograms of seed voxels as a distribution of points in 3D MNI space using GMMs as introduced in Chapter 2. GMMs are parametric probability density functions that have been shown to be useful in approximating densities with complicated shapes. Figure.2.2 shows 3 randomly selected tractograms of the seed voxels, and their corresponding Gaussian mixture models.

Experiments have shown that setting the number of components to between 15 and 20 is sufficient to characterize the shape of a tractogram, little difference in application can be seen within this range. We picked 17 as a usually used number of mixtures in this study. The red dots in the bottom pictures in Figure 2.2 are the mean of each mixture of the GMMs.

We take advantage of *gmdistribution,* a powerful class from the MATLAB © Statistics Toolbox, to fit the GMMs. *Gmdistribution* implements the EM algorithm in an efficient and robust way. Via setting the 'Replicates' option within *gmdistribution,* the fitting of every voxel is repeated for 5 times and the result with the largest likelihood is returned in order to avoid falling into local optimum as possible. Initial means of mixtures were uniformly randomly selected from the tractograms for initial parameter estimation.

5.4 Connectivity-based Parcellation

By employing GMMs, tractograms are parameterized and smoothed. We used the earth mover's distance (EMD, Rubner, 2000) to compute the distance between to tractograms. The EMD corresponds to minimize the cost of moving 'material' from one distribution to another, and is computed using linear programming methods. In the EMD's point of view, a distribution can be represented by a histogram where each bar is represented by its mean (or mode), and by the fraction of the distribution that belongs to that bar. We call such a representation the *signature* of the distribution.

Given two signatures $P = \{(p_i, u_i)\}_{i=1}^{m}$ and $Q = \{(q_j, v_j)\}_{j=1}^{n}$ with size *m, n* respectively, the EMD between *P* and *Q* is modeled as a solution to a transportation problem. Treat *P* as "supplies" located at u_i and elements in Q as "demands" at v_i. Then p_i and q_j indicates the amount of supply and demand respectively. The EMD is defined as the minimum (normalized work required for resolving the supply-demand transports:

$$EMD(P,Q) = \min_{F=\{f_{ij}\}} \frac{\sum_{i,j} f_{ij} d_{ij}}{\sum_{i,j} f_{ij}}$$

$$s.t. \begin{cases} \sum_j f_{ij} \le p_i, \\ \sum_i f_{ij} \le q_j, \\ \sum_{i,j} f_{ij} = \min\{\sum_i p_i, \sum_j q_j\}, f_{ij} \ge 0. \end{cases} \quad (5.4.1)$$

where $F = \{f_{ij}\}$ denotes a set of *flow,* representing the amount transported from the *i-th* supply to the *j-th* demand. d_{ij} is called the ground distance between position u_i and v_j. With a simplified solution to the classic transportation problem, the complexity of computing EMD can be reduced to $N^2 \log^2 N$, where N denotes the scale of the number of signatures (Pele & Werman, 2009).

The distributions were sampled from the GMM of the tractogram rather than the tractogram itself. The sampling strategy works as follows: first, deciding the number of samples, we picked 4 multiplication of the number of mixtures, in this case $4 \times 17 = 68$ signatures; second, sampling the signatures from the GMM using importance sampling; finally, calculating the ground distances by the Euclidean distances between each pair of signatures. A distance matrix of one subject's inferior parietal lobule is shown in Figure 3.2 (a).

The distance (dissimilarity) measures d_{ij} then needs to be transformed to similarity measures s_{ij} for clustering use. I took the exponential transformation:

$$s_{ij} = \frac{1}{2}\exp[-\frac{(d_{ij}-u)^2}{2\sigma^2}] \tag{5.4.2}$$

in which u and σ denotes the mean and standard deviation of the distances respectively.

So far, we have got the similarity matrix of the tractograms. Distance-based clustering is to be applied to start the connectivity-based parcellation. We employed two clustering schemes to conduct CBP, each for a different research purpose.

Our first problem is the multi-subject CBP of the inferior parietal lobule (IPL). Atlhough the connectional structure of IPL is highly heterogeneous, the number of parcels is believed (on cytoarchitectonic grounds) to be within the range from 5 to 7 (Caspers, et al., 2008). According to Juelich Atlas (Eickhoff, et al. 2005), there are 7 sub-regions within the IPL area of each hemisphere of the brain. So 7 clusters are used in our model. Using a predefined number of clusters, spectral clustering was applied to provide an initial partition of the IPL voxels. A clustered distance matrix with spectral clustering is displayed in Figure 3.1.2(b). After the initialization, a hidden Markov random field (HMRF) model is used to refine the clustering. We treat the sub-region label of the seed voxels as the hidden variables and the connectivity profiles as the observations. In the HMRF framework, a hidden variable is spatially constrained by its neighbors, i.e. adjacent voxels are expected to have similar labeling. To accomplish this, a Markov random field (MRF) is used to model local contextual information. As described in Chapter 3, an iterated method is used to infer the hidden variables of the HMRF. For the E-step, an exemplar tractogram of each subregion is selected. The exemplar represents an "average" connectivity pattern for its sub-region, which was obtained by averaging the tractograms

of the voxels with same labels. For the M-step, voxels are re-labeled according to the maximum a posteriori of the MRF given the current exemplars. The E-step and M-step were alternated until the labeling stopped changing significantly.

Then we investigated whole brain CBP. To parcellate an arbitrary region of human cortex, the number of sub-regions is not well-defined. We thus need to use non-parametric Bayes models to learn the number from data. We employed a spectral Chinese restaurant process model introduced by Socher (2011). The approach combines non-parametric Chinese restaurant process (CRP) and spectral methods for dimensionality reduction. This scheme uses spectral clustering to project the data from a high dimensional connectivity space to a lower dimensional Euclidean space (2-D or 3-D); then clusters the voxels in this lower dimensional space with Dirichlet process model (DPM). For the whole brain CBP, we focused on inferior / superior parietal lobules (IPL/SPL), and precentral / post central gyri.

5.5 Multi-subject CBP

For the study of IPL, after each individual subject was parcellated, subjects were merged, combining the tractograms of all subjects at each point in MNI space. The MNI is a standard brain coordinate system defined a large series of MRI scans on normal controls (Mazziotta, et al., 1995).

To create a tractogram for each seed voxel, we uniformly sampled from the tractograms from the corresponding voxels in the individual datasets. We then performed CBP on the merged voxels to create a joint parcellation map. The map was generated by overlapping the seed voxels of each individual subject's IPL area and merging on a maximum probability basis (Tungaraza, et al., 2012).

Finally, the joint parcellation map is used in HMRF-based CBP. Exemplars of each parcel of the atlas are extracted and the voxels are grouped based on their similarity in tractograms to the exemplars, *i.e.,* the voxels are assigned the label of the closest exemplar. An MRF model is finally used to denoise the segmentation. To study the reproducibility of the atlas, we created joint atlases from brains 2-8 and 9-16 and re-parcellated brain 1 using each of the joint atlases.

5.6 Resolving Crossing Fibers With Compressed Sensing

The experiments of crossing fibers resolution consist two parts: simulation and empirical data.

Simulation. Two fiber tracts comprising tensors having FA=0.7 ($\lambda_1 = 2\times10^{-3}\text{mm}^2/\text{s}$, $\lambda_2=\lambda_3=0.5\times10^{-3}\text{mm}^2/\text{s}$) were simulated such that they cross at a certain angle. A typical clinical DTI

sequence with b-value of 1000 s/mm^2, and Rician noise at SNR of 25:1 was simulated. Based on Eq (4.2.1), a set of basis was built with high angular resolution to uniformly cover the hemisphere, see Figure 5.1. A voxel with crossing fibers is supposed to be constructed by a linear combination of vectors from the basis. The vectors are shuffled in the basis such that the transform matrix is incoherent in this domain. The optimization problem in (4.2.3) is solved by the convex optimization MATLAB package *CVX* (http://cvxr.com/cvx/) developed by Michael Grant and Stephen Boyd.

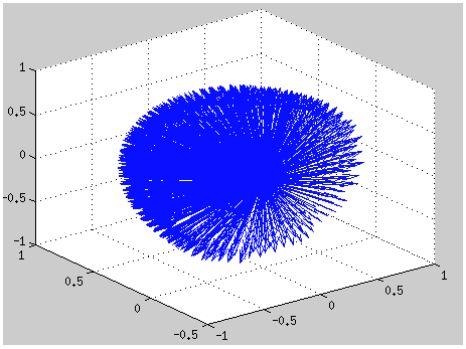


Figure 5.1. 1024 bases vectors uniformly distributing over a hemisphere.

Experiments were conducted on various numbers of measurements (M=16~128), basis vector resolution (N= 128~1024), and crossing angle (0°~ 90°). The performance of the simulation was evaluated by a cone of uncertainty error metric:

- Given a simulated set of crossing fibers $R:\{t_j,w_j\}_{j=1}^{M}$ and the estimate set $E:\{f_i,v_i\}_{i=1}^{N}$, where v_i, w_j are the orientation of the vectors and f_i, t_j are the weights of corresponding vectors, The fundamental cone is defined by the collection of estimated directions closest to a reference direction whose total estimated partial fraction do not exceed the reference fraction

$$\theta_j = \sum_{k=1}^{K_j+1} f_{k,j}^{i} \angle(v_{k,j}, w_j) \tag{5.6.1}$$

  $v^i_{k,j}$ denotes the k^{th} closest estimated vector to reference vector w_j.

- The estimate's cone of uncertainty to the reference is written as the squared weighted sum of angular error

$$\psi(E,R) = \frac{\sum_j t_j^2 \theta_j}{\sum_j t_j^2} \tag{5.6.2}$$

- To make the metric symmetric

$$\psi = \frac{\psi(E,R) + \psi(R,E)}{2} \tag{5.6.3}$$

Considering the randomness in the solution of the CS problems, each simulation was repeated for 50 times, such that an average angular error could be obtained.

Empirical Data. The reconstruction basis was built as for the simulation. The CS based crossing fibers resolution approach was applied to the DWI volumes obtained with 128 gradients and b-value = 1000s/mm^2. Each voxel in the DWI would have a vector containing the orientations and fractions of the non-zero bases. A probabilistic tractography algorithm was implemented and imported to the MRtrix framework. The CS based tractograms were compared to that obtained from constrained spherical deconvolution (CSD) methods both visually and by their CBP results.

Chapter 6 RESULTS AND ANALYSIS

The results in this chapter are focused on the multi-subject CBP of IPL, single-subject CBP of SPL/SPL and pre/post central, and crossing fibers resolution using CS.

6.1 Multi-subject CBP of IPL

The spectral clustering-HMRF parcellation pipeline was applied to the 16 subjects (see Figure. 6.1). The IPL areas are divided into 7 sub-regions. The 7 parcels are basically vertically aligned, which is consistent with the cytoarchitectonic literature (Caspers, et al., 2008; Tungaraza, et al., 2012). Our approach to CBP gave partitions on IPL that were consistent across subjects.

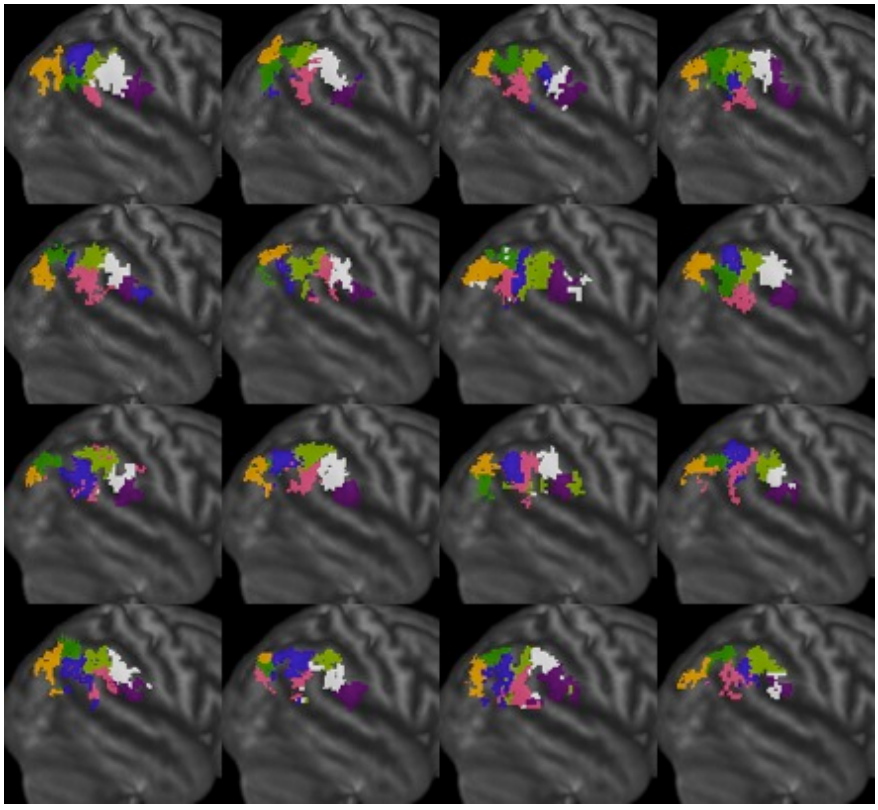


Figure. 6.1. Single-subject CBP on 16 subjects (right hemisphere); area corresponds to yellow rectangle in Figure. 6.2.

We then performed HMRF parcellation jointly on all 16 subjects, as described in section 5.5. The joint segmentation is compared to a cytoarchitectonic atlas of IPL in Figure 6.2. The map was generated on a MAP basis. The consistency of the joint parcellation with the Juelich atlas was evaluated by the Dice coefficient. An average Dice coefficient of 65% (69% on the left hemisphere and 61% on the right hemisphere, respectively, Table 6.1 was observed.

Table 6.1 Dice Coefficients of Single/Multi-subject CBP (%)

	Single-subject CBP			Multi-subject CBP		
Subjects	Left Hemi	Right Hemi	Average	Left Hemi	Right Hemi	Average
1	72.5	58.3	65.4	70.5	64.6	67.6
2	67.5	60.1	63.8	73.8	67.5	70.7
3	65.3	60.5	62.9	70.6	64.1	67.4
4	66.9	63.5	65.2	70.1	64.1	67.1
5	69.1	62.8	65.9	72.2	66.0	69.1
6	67.5	63.3	65.4	72.5	67.6	70.0
7	72.9	62.4	67.7	70.6	66.9	68.7
8	69.6	60.3	64.9	72.0	64.5	68.3
9	68.3	63.4	65.9	71.2	64.4	67.8
10	71.7	60.6	66.2	70.1	65.4	67.7
11	69.0	60.2	64.6	70.6	67.4	69.0
12	70.5	61.2	65.9	72.4	65.1	68.7
13	70.3	61.6	65.9	72.0	66.0	69.0
14	70.5	61.1	65.8	71.2	65.8	68.5
15	66.7	60.8	63.7	70.5	64.7	67.6
16	72.1	62.7	67.4	73.6	65.5	69.6
Mean	69.4	61.4	65.4	71.5	65.6	68.6

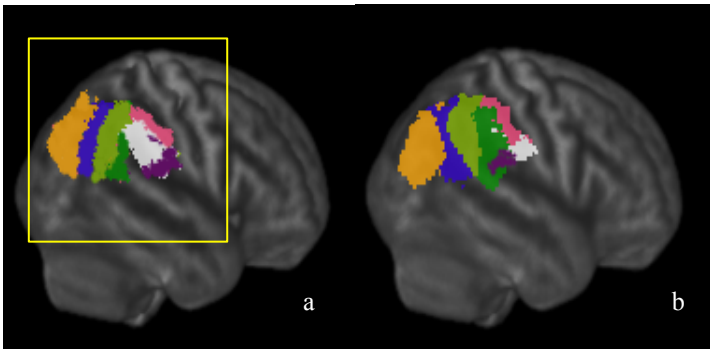


Figure 6.2. Multi-subject parcellation of IPL (right hemisphere). (a) HMRF parcellation; (b) FSL Juelich atlas. Yellow rectangle indicates area of Figure. 6.1.

Exemplars were extracted from each sub-region of the merged map. Classifiecation based on the joint exemplars was conducted on the 16 individual subjects. The new parcellations (see Figure. 6.3) appear generally smoother and more consistent with each other and with the Juelich atlas than the results of single-subject CBP (Figure.6.2).

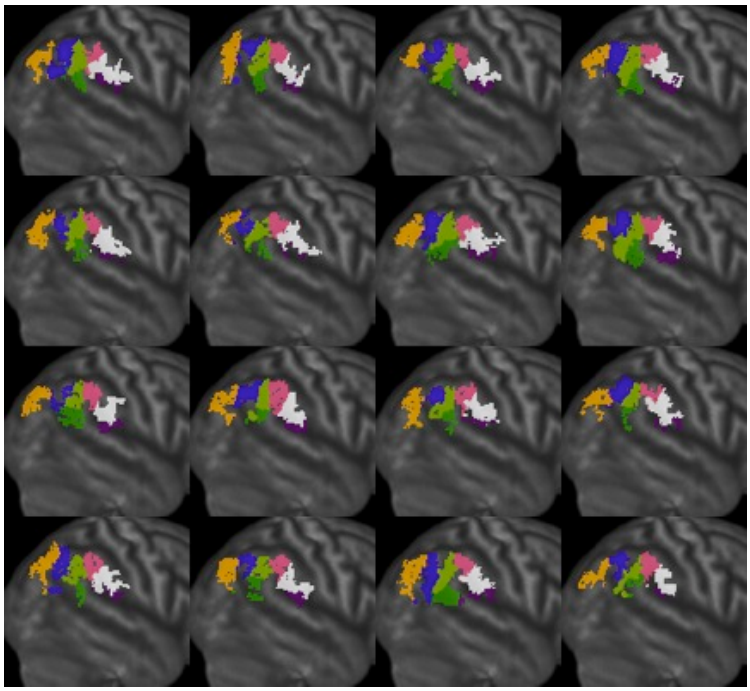


Figure.6.3. CBP on 16 individual subjects using exemplars extracted from the merged CBP atlas. The configurations of the segmentation shows high resemblance with the Juelich atlas. Compare with Figure. 6.2, in which each subject was parcellated separately.

To examine the effects of the joint atlas on the individual parcellations, we separately parcellated brain 1 using brains 2-8 and 9-16, respectively, to make the joint atlas. Figure 6.4 shows that the results are very similar (average Dice coefficient, 69%), suggesting that the parcellation results are relatively independent of the brains used to compute the joint atlas.

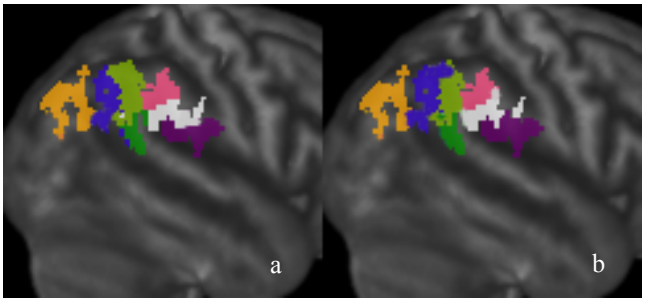


Figure. 6.4. CBPs of brain 1 using joint atlas generated from brains 2-8 (a) and 9-16 (b), respectively, showing that the CBP results are relatively independent of the subjects chosen to create the atlas.

6.2 CBP With Non-parametric Model

In this section, we learn the number of parcels within areas of interest. We examine the impact of the hyper-parameters on the number of parcels and the driven force in connectional structure segmenting the cortical regions.

When I tried to extend CBP on IPL to the whole brain CBP, I found that most of the cortical regions, such as the SPL and the pre/postcentral gyri, are not so heterogeneous as IPL. Consequently some aspects of the CBP need to be modified, including the strategy of generating tractographies. Among the options of tracks generation, whether the tracks grow unidirectionally or bidirectionally matters. By default, MRtrix generates probabilistic tractography in two directions from each seed voxel, which results in a dominant cluster of the end points of tracks around their seed voxel, leading to troubles in the fitting GMMs and calculating the distance between tractograms, see Figure 6.5(a). The new tractographies were generated with a *–unidirectional* command. The artifacts of the dominant near-seed cluster is ruled out and the true parcellation revealed (Figure 6.5(b)). According to the motor homunculus representation (Marie & Hoehn, 2009) in primary motor cortex, primary motor cortex (M1), premotor cortex (PMA) and supplementary motor cortex (SMA), is responsible for the planning and executing body movements. And the functionality partition is aligned horizontally, from top to bottom, controlling lower limbs, upper limbs and face respectively. The primary sensory cortex (postcentral) has the corresponding anatomical divisions with sensory functionalities. Thus an ideal CBP of the pre/post centrals is supposed to contain 6 sub-regions, with 3 for precentral and postcentral each. Figure 6.5(b) is obtained with the *–unidirectional* option, which generates the tracks in one direction. In the segmentation obtained by bi-directional tractographies (see Figure 6.5(a)), one can hardly find the boundary between precentral and postcentral.

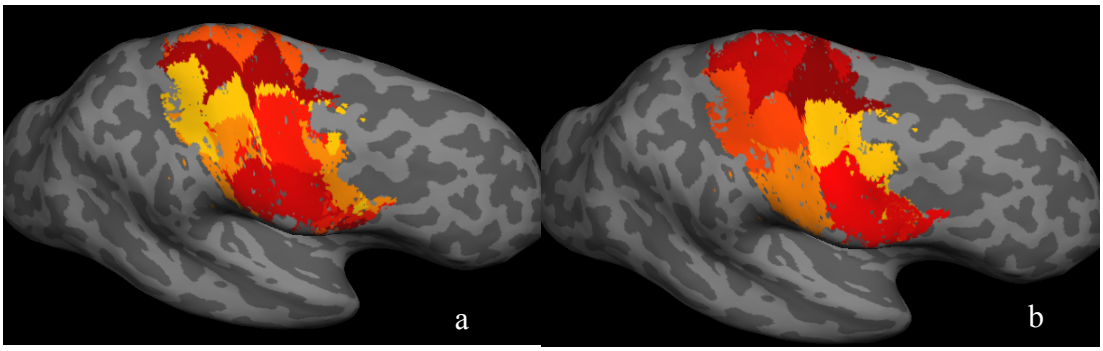


Figure 6.5 (a) CBP using bi-directional tracks (b) CBP using uni-directional tacks

The hyper parameters of DPM, such as the concentration parameter α, also play a crucial role in the clustering. Generally speaking, the larger α is, the more clusters will be found. In my implementation, α is set to be a fraction of the size of the data *n, i.e.* $\alpha = \frac{n}{b}$. We tested the b in the range of 5~5000 (default is 50), larger *b* implies smaller α, thus fewer clusters. The results are shown in Figure 6.6. As *b* decreases, α increases, and the number of clusters increases. The hyper parameter provides reasonable results in the range $\alpha = \frac{n}{500} \sim \frac{n}{50}$.

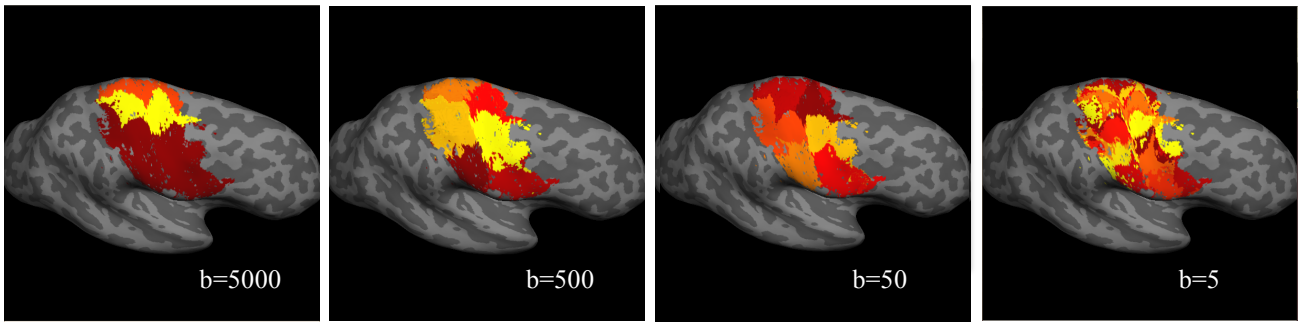


Figure 6.6. The impact of the concentration parameter $\alpha = n / b$ on the number of clusters *c*. From left to right: *b=5000, c=3; b=500, c=6; b=50, c=6; b=5, c=18*

We then compare the boundaries of anatomical segmentation given by the FSL atlas and our connectional segmentation.

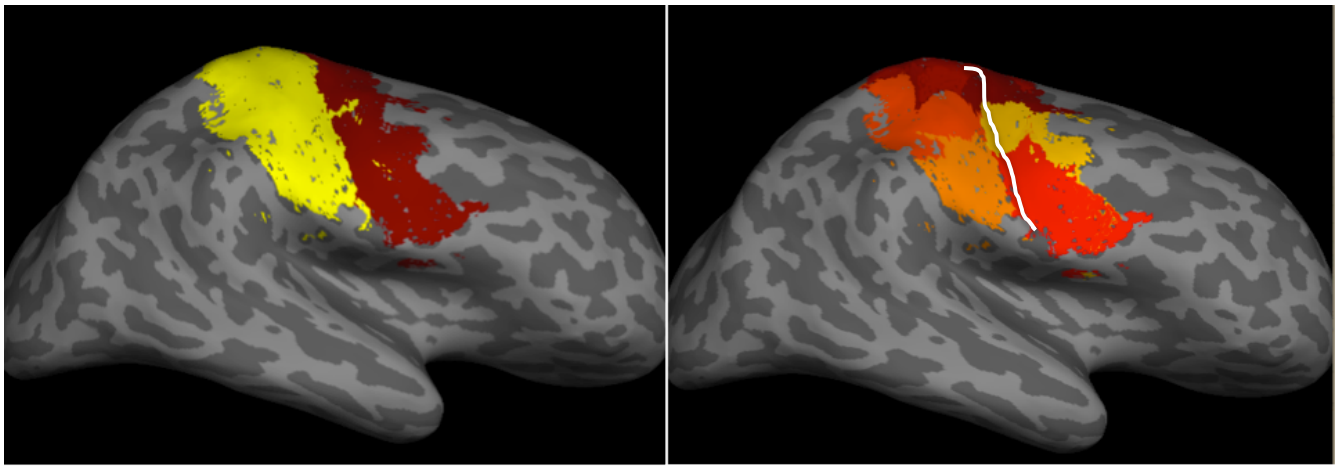


Figure 6.7 (left) FSL segmentation of the pre/post centrals. (right) CBP of the pre/post centrals. The white line shows the pre/post centrals boundary of FSL segmentation

Figure 6.7 compares CBP to the FSL segmentation in the pre/post central areas. For CBP, precentral and postcentral each is divided into 3 sub-regions, horizontally aligned. The boundary between precentral and postcentral is obvious. The boundary is consistent with that of the FSL segmentation, except for that the boundary shifts slightly to left in CBP compared to FSL.

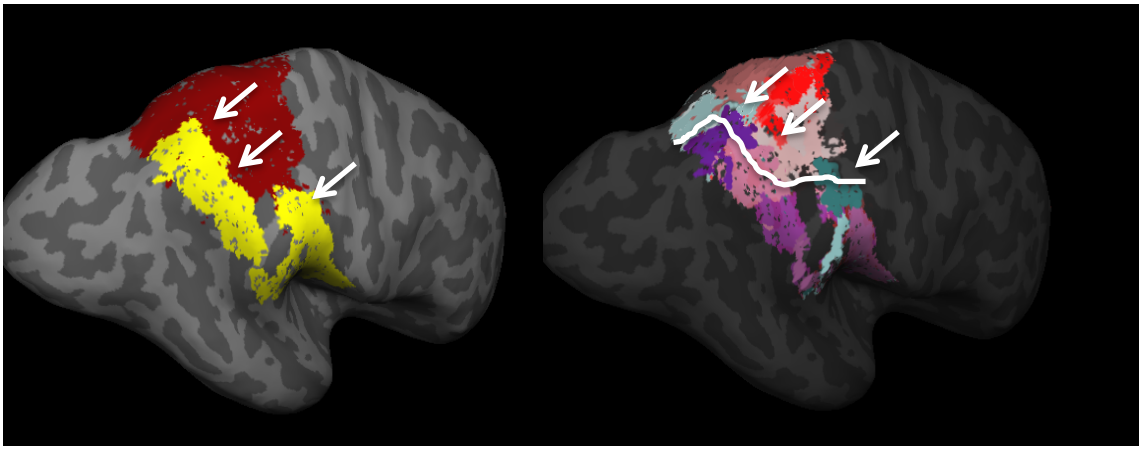


Figure 6.8 (left) FSL segmentation of superior parietal lobule (SPL, yellow) and IPL (brown). (right) CBP of SPL and IPL. The white line shows the pre/post centrals boundary of FSL segmentation The arrows mark the major disagreements between two

Figure 6.8 compares CBP to the FSL segmentation in the SPL and IPL areas. The non-parametric CBP pipeline parcellates the SPL into 5 sub-regions and IPL 6 sub-regions (unlike we did using spectral clustering). The boundaries are generally similar besides some disagreements which are marked by the arrows.

6.3 Resolving The Crossing Fibers Regions Using Compressed Sensing

Simulations. We first fixed the angle between two crossing fibers ($K=2$) at 90°, and tested the impact of the size of the reconstruction basis with a) 48 measurements and b) 128 measurements. Larger basis size suggests higher spatial resolution of the basis vectors. Figure 6.9 shows that the angular error between simulating data and estimation is reduced with the increase of N (from 128 to 1024). With fewer measurements, the mean error and the standard deviation will get bigger.

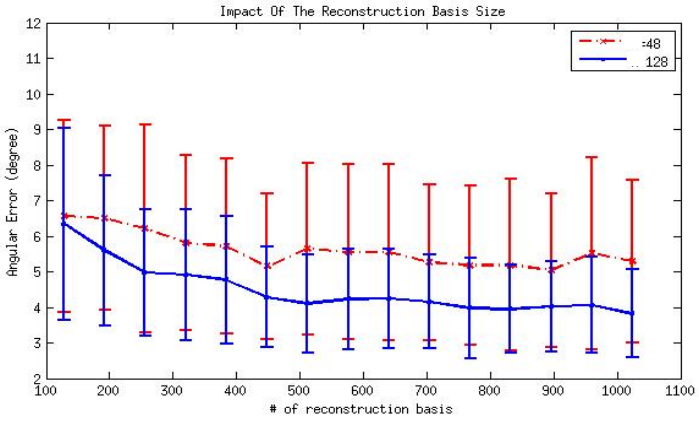


Figure. 6.9 Impact of reconstruction resolution (number of basis *N*). (red)*M*=48; (blue) *M*=128.

We then investigated the impact of the number of measurements *M* (Figure 6.10) and the crossing angle (Figure 6.11) on the accuracy of the reconstruction. The number of measurements varies from 16 to 128 with a fixed *N=1024*. The both the mean and the standard deviation of the angular error decreases as the *M* grows. While for $M > 20$, the decrease slows down. Consider the requirement of the measurement number in section 4.1,

$$M \geq K \log_2(N/K) = 2 * \log_2(1024/2) = 18 \ ,$$

which agrees with the results in (Figure 6.10).

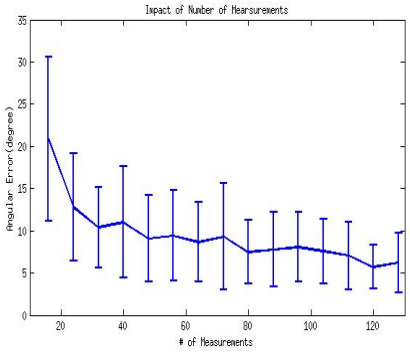


Figure 6.10. Impact of number of measurements (number of gradients *M*). Angular error decreases as *M* grows

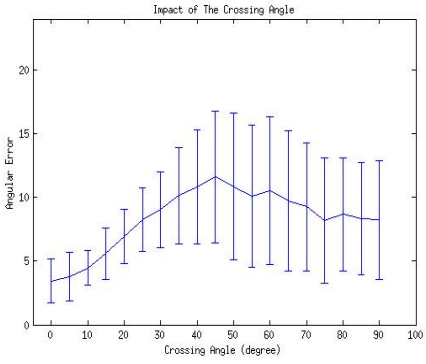


Figure 6.11. Impact of crossing angle. The angular error first increases then drops as the crossing angle changes from 0° to 90°.

In Figure 6.11, the angle between two crossing fibers ranges from 0° to 90°. The error first increases then drops. The intuitive explanation is that it is easier to resolve to fibers when the crossing angle is either small or big. When the angle is small (~0°), they can be treated as a single fiber; whereas large

angles (~90°) have a higher tolerance for the spatial resolution of the basis.

Empirical Data.

Figure. 6.12 displays examples of resolved crossing fibers within single voxels. Figure 6.12(a1) indicates a typical region of crossing fibers at the junction of corpus callosum and white matter. A random voxel was selected from the yellow rectangular region to show in Figure 6.12(a) and other 4 voxels in Figure 6.12(b). Figure 6.12(a) compares the resolution results between constrained spherical deconvoluiton (CSD) given by MRtrix and those obtained by CS with different numbers of measurements (a3, M=64; a4, M=32). Both CS reconstructions appear similar to the CSD result; the estimate with 64 gradients looks slightly shaper than that with 32 gradients.

The top pictures in Figure 6.12(b) show 4 voxels resolved by CSD and the bottom ones show the corresponding CS estimates with 64 gradients. Though minor differences in shape and orientations can be observed, results obtained by compressed sensing approach are highly consistent with those by traditional methods, and with lower cost.

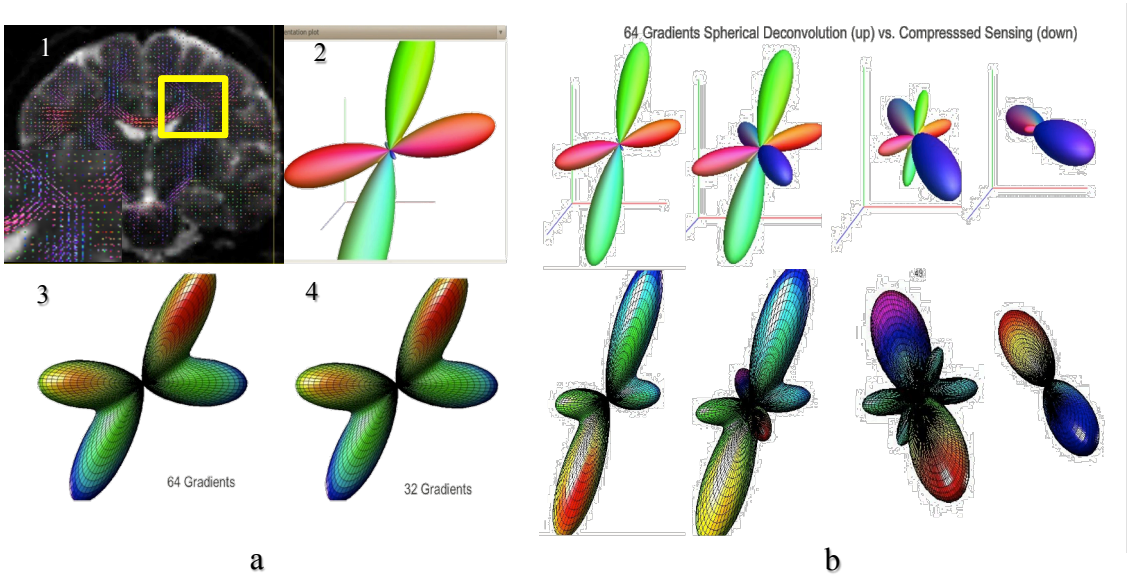


Figure 6.12. (a1) a region of crossing fibers; (a2) voxel resolved using CSD with 128 measurements; (A3) same voxel resolved using CS with 64 measurements; (a4) voxel resolved using CS with 32 measurements. (b) Examples of other 4 voxels with different orientations and # of fiber crossed: (top) by CSD; (bottom) by CS.

We also applied the compressed sensing approach to tractography. 5000 tracks were generated from one seed voxel in each scenario. Results acquired with different numbers of measurements (*M=32,48, 64*) are compared to the tractography generated with CSD with 128 measurements in Figure 6.13. For the cases *M*=48 and 64, the more tracks and details were present than *M=32.* No significant difference can be observed in the *M*=48 and 64 cases, compared to CSD results.

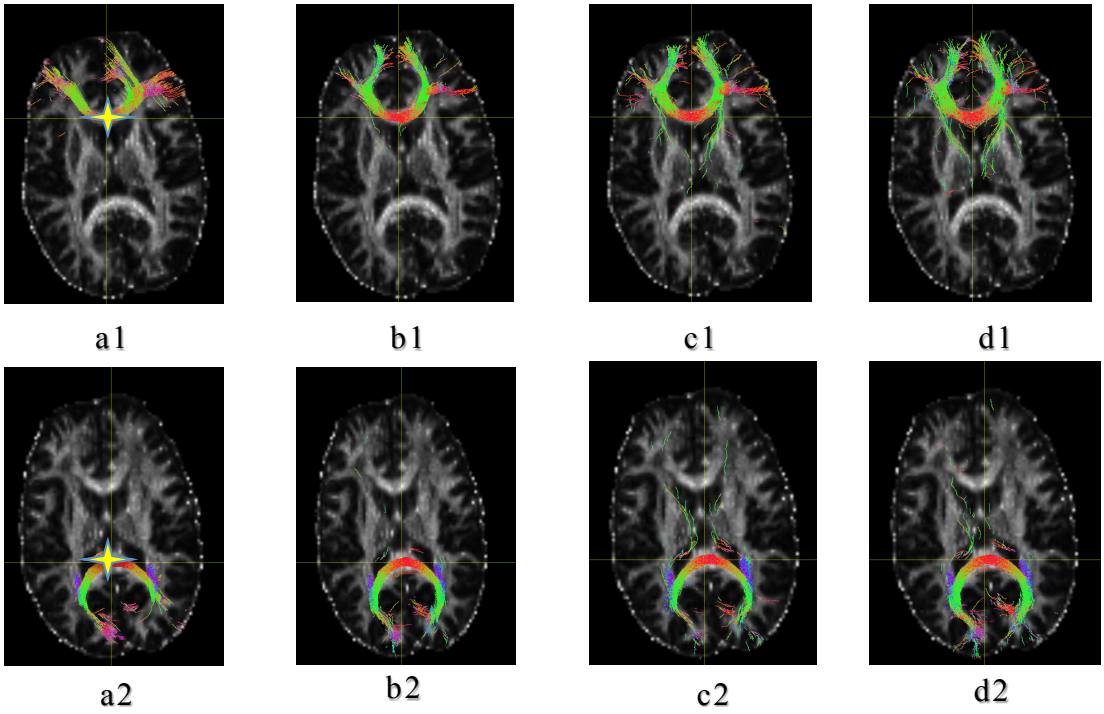


Figure 6.13. Two sets of tractography results with different numbers of measurements (a) *K*=32 (CS) (b) *K*=48(CS) (c) *K*=64(CS) (d) *K*=128(CSD). Two seeds (yellow cross in the top and bottom graphs respectively) ware planted within corpus callosum. A probabilistic framework of tractography was conducted . The tracks graphs were overlapped on fraction anisotropy maps.

The CS-based tractography was further tested with the CBP process. We follow the multi-subject CBP pipeline described in section 5.5 with probabilistic tractography generated using CS. A new CBP atlas for IPL was built and compared to the former results. Figure 6.14(c) shows the atlas obtained from the CS-based CBP; Figure 6.14(a,b) are the same in Figure 6.2. The Dice coefficients of CS-based map is 63% in average, 60% in the right hemisphere and 66% in left hemisphere.

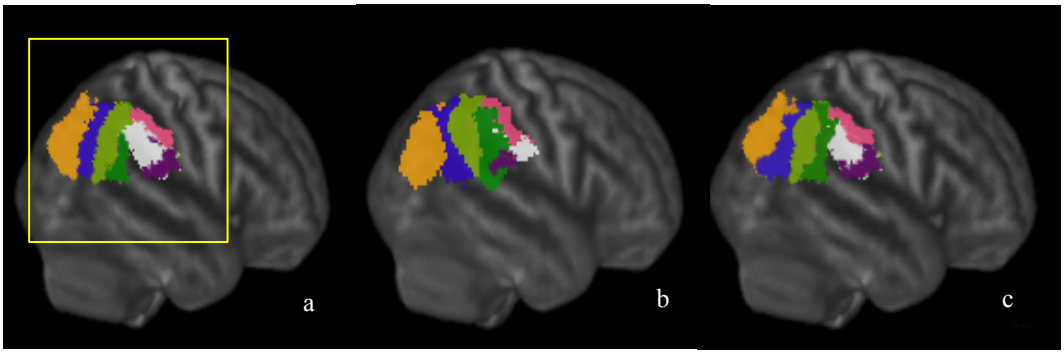


Figure. 6.14. (a) CBP atlas using CSD generated tractograms; (b) FSL Juelich atlas; (c) CBP atlas using CS generated tractograms

Chapter 7 CONCLUSION AND DISCUSSION

A multi-subject connectivity-based parcellation pipeline is presented via a new GMM tractogram in this thesis. The research may benefit both neuroscience and clinical practice.

7.1 Achievements

In the study of multi-subject CBP of IPL, we described a GMM-HMRF approach to connectivity-based parcellation. By employing this methodology, we defined a new description of connectivity profile (GMM); also we were able to segment human IPL areas taking into consideration the spatial interaction of adjacent voxels (HMRF). We demonstrated a high similarity between our partition and the standard cytoarchitectonic map. Moreover, a CBP atlas was built by merging subjects' parcellation. Using the exemplars extracted from the CBP atlas, both the computational complexity and segmentation consistency of new parcellation are improved. A multi-subject CBP approach provides a non-invasive method to explore human cortical connectivity, and should significantly enhance our understanding of normal and abnormal brain function.

We then extended our CBP framework to the other cortical areas, including pre/post centrals and SPL, by combining the spectral methods with non-parametric Bayes models. The spectral non-Bayes models, Dirichlet process model in particular, enables the data learn the number of clusters automatically. Cluster number is significantly controlled by the hyper-parameter , the clusters become fewer when decreases. We found, however, that the choose of is flexible to some extend, since the applicable range is relatively large: *n/500~n/50,* where n is the size of dataset. By investigating the CBP of pre/post central areas, we found that the connectional factor driving the parcellation consents with the motor homunculus model, which implies that the consistency between connectivity and functionality.

The general agreement between FSL segmentation, which is obtained from a cytoarchitectonics-based atlas, and the connectivity-based parcellation also verifies this argument.

In the research of crossing fiber resolution, the compressed sensing technique shows great promise in resolving crossing fiber issues. With compressed sensing, the number of measurements can be reduced to 32 or 48 (48 is more favorable in the this study) with higher efficiency and moderate accuracy. It works well with small or big crossing angles. Compressed sensing results can be applied to tractography with a probabilistic or deterministic framework. The experiments combined with CBP demonstrate its availability in application.

7.2 Disadvantages

The research has the following weaknesses. First, GMMs are sensitive to outliers. Some tractography with very low probability may exert a vital influence on the fitting of the models. Our method is not very robust for such outliers. Second, the spatial resolution of GMM is relatively low and easy to be affected by noise. For example, the thalamic connections are crucial for identifying the connection pattern of each cortical segment, but due to its small size, thalamus doesn't earn enough credits in GMM-based CBP. Third, although mathematically elegant, spectral clustering doesn't apply to large scale clustering problems (like whole brain CBP), since it requires matrix decomposition. Other derivative spectral techniques, such as graph cuts may be worth considering.

7.3 Future Work

Our methods have proved its potential in connectivity-based parcellation. The research in future may follow several directions:

1. Extending to the whole brain CBP. Due to the high variation in the connectional characteristics over different cortical regions, our pipelines can be currently applied to certain regions locally, which limits its use in the whole brain study. Prospective solutions are either finding global approaches, or using a predefined segmentation as *a prior* and iteratively updating the segments using Bayes rule.
2. An atlas of whole brain CBP is to be constructed. Therefore future CBP can be performed by registratered to the atlas. Also, developing atlas (from infants to adults) and disease specific atlas are interesting directions of research.
3. Taking advantage of more information from tractography. Our representation of connectivity patterns by the distribution of end points has left much important information carried by the entire tracks, such as the length, the pathway, and connections along the tracks. For example, many connections between brain segments are "passing-by" rather than "ending-at"; our methods are inadequate in characterizing the former situation. A new representation is expected to take account of all the 3D information of tractography.
4. Combining with the functional studies. It's presumed that the structural partitions are consistent with regional functionalities. Finding the mapping between connectional and functional segmentations can also verify the conclusion of both studies.
5. Refining the CS-based crossing fiber resolution, so that scanning time can be significantly reduced.

BIBLIOGRAPHY

[1] Anderson, A. W. (2001). Theoretical analysis of the effects of noise on diffusion tensor imaging. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *46*(6), 1174–88. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11746585

[2] Andrieu, C., de Freitas, N., Doucet, A., & Jordan, M. I. (2003). An Introduction to MCMC for Machine Learning. *Machine Learning*, *50*(1-2), 5–43. doi:10.1023/A:1020281327116

[3] Anjari, M., Srinivasan, L., Allsop, J. M., Hajnal, J. V., Rutherford, M. a, Edwards, a D., & Counsell, S. J. (2007). Diffusion tensor imaging with tract-based spatial statistics reveals local white matter abnormalities in preterm infants. *NeuroImage*, *35*(3), 1021–7. doi:10.1016/j.neuroimage.2007.01.035

[4] Anwander, a, Tittgemeyer, M., von Cramon, D. Y., Friederici, a D., & Knösche, T. R. (2007). Connectivity-Based Parcellation of Broca's Area. *Cerebral cortex (New York, N.Y. : 1991)*, *17*(4), 816–25. doi:10.1093/cercor/bhk034

[5] Baraniuk, R. (2007). Compressive Sensing [Lecture Notes]. *IEEE Signal Processing Magazine*, *24*(4), 118–121. doi:10.1109/MSP.2007.4286571

[6] Basser, P. J., Mattiello, J., & LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, *66*(1), 259–67. doi:10.1016/S0006-3495(94)80775-1

[7] Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., & Aldroubi, a. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *44*(4), 625–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11025519

[8] Beckmann, M., Johansen-Berg, H., & Rushworth, M. F. S. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *29*(4), 1175–90. doi:10.1523/JNEUROSCI.3328-08.2009

[9] Behrens, T. E. J., Johansen-Berg, H., Woolrich, M. W., Smith, S. M., Wheeler-Kingshott, C. a M., Boulby, P. a, Barker, G. J., et al. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature neuroscience*, *6*(7), 750–7. doi:10.1038/nn1075

[1] Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, *B-48*, 259 – 302. Retrieved from http://www.citeulike.org/user/christianwolf/article/2335709

[2] Besag, Julian. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems Julian Besag. *Journal of the Royal Statistical Society*, *36*(2), 192–236.

[3] Bishop, C. M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)* (p. 738). Springer. Retrieved from http://www.amazon.com/Pattern-Recognition-Learning-Information-Statistics/dp/0387310738

[4] Bruckstein, A. M., Donoho, D. L., & Elad, M. (2009). From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Review*, *51*(1), 34–81. doi:10.1137/060657704

[5] Candes, E. J., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, *52*(2), 489–509. doi:10.1109/TIT.2005.862083

[6] Caspers, S., Eickhoff, S. B., Geyer, S., Scheperjans, F., Mohlberg, H., Zilles, K., & Amunts, K. (2008). The human inferior parietal lobule in stereotaxic space. *Brain structure & function*, *212*(6), 481–95. doi:10.1007/s00429-008-0195-z

[7] Caspers, S., Eickhoff, S. B., Rick, T., von Kapri, A., Kuhlen, T., Huang, R., Shah, N. J., et al. (2011). Probabilistic fibre tract analysis of cytoarchitectonically defined human inferior parietal lobule areas reveals similarities to macaques. *NeuroImage*, *58*(2), 362–80. doi:10.1016/j.neuroimage.2011.06.027

[8] Catani, M., & Schotten, M. T. de. (2012). *Atlas of Human Brain Connections* (p. 456). Oxford University Press, USA. Retrieved from http://www.amazon.com/Atlas-Human-Brain-Connections-Catani/dp/0199541167

[9] Chung, F. R. K. (1997). *Spectral Graph Theory, Issue 92 (Google eBook)* (p. 207). American Mathematical Soc. Retrieved from http://books.google.com/books/about/Spectral_Graph_Theory.html?id=YUc38_MCuhAC&pgis=1

[10] Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–94. doi:10.1006/nimg.1998.0395

[11] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.

[12] Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, *52*(4), 1289–1306. doi:10.1109/TIT.2006.871582

[13] Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, *1*(2), 209–230. Retrieved from http://projecteuclid.org/euclid.aos/1176342360

[14] Frank, L. R. (2002). Characterization of anisotropy in high angular resolution diffusion-weighted MRI. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *47*(6), 1083–99. doi:10.1002/mrm.10156

[15] Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(6), 721–741. doi:10.1109/TPAMI.1984.4767596

[16] Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*(1), 1–12. doi:10.1016/j.jmp.2011.08.004

[17] Gotlib, I. H., Sivers, H., Gabrieli, J. D. E., Whitfield-Gabrieli, S., Goldin, P., Minor, K. L., & Canli, T. (2005). Subgenual anterior cingulate activation to valenced emotional stimuli in major depression. *Neuroreport*, *16*(16), 1731–4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16237317

[18] Hahn, G., & Sabidussi, G. (Eds.). (1997). *Graph Symmetry*. Dordrecht: Springer Netherlands. doi:10.1007/978-94-015-8937-6

[19] Hosey, T., Williams, G., & Ansorge, R. (2005). Inference of multiple fiber orientations in high angular resolution diffusion imaging. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *54*(6), 1480–9. doi:10.1002/mrm.20723

[20] Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Drobnjak, I., Rushworth, M. F. S., Brady, J. M., Smith, S. M., et al. (2004). Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13335–40. doi:10.1073/pnas.0403743101

[21] Kim, S.-J., Koh, K., Lustig, M., & Boyd, S. (2007). An Efficient Method for Compressed Sensing. *2007 IEEE International Conference on Image Processing* (Vol. 3, p. III – 117–III – 120). IEEE. doi:10.1109/ICIP.2007.4379260

[22] Kim, Seung-Jean and Boyd, S. (n.d.). AN EFFICIENT METHOD FOR COMPRESSED SENSING, (3), 3–6.

[23] Kubicki, M., McCarley, R., Westin, C.-F., Park, H.-J., Maier, S., Kikinis, R., Jolesz, F. a, et al. (2007). A review of diffusion tensor imaging studies in schizophrenia. *Journal of psychiatric research*, *41*(1-2), 15–30. doi:10.1016/j.jpsychires.2005.05.005

[24] Landman, B. a, Bogovic, J. a, Wan, H., El Zahraa ElShahaby, F., Bazin, P.-L., & Prince, J. L. (2012). Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI. *NeuroImage*, *59*(3), 2175–86. doi:10.1016/j.neuroimage.2011.10.011

[25] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., & Laval-Jeantet, M. (1986). MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, *161*(2), 401–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3763909

[26] Li, S. Z. (1995). *Markov Random Field Modeling in Computer Vision* (p. 264). Springer-Verlag. Retrieved from http://books.google.com/books?id=k1R0QgAACAAJ&pgis=1

[27] Lustig, M., Donoho, D. L., Santos, J. M., & Pauly, J. M. (2008). Compressed Sensing MRI. *IEEE Signal Processing Magazine*, *25*(2), 72–82. doi:10.1109/MSP.2007.914728

[28] Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416. doi:10.1007/s11222-007-9033-z

[29] Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 888–905. doi:10.1109/34.868688

[30] Marieb, E. N., & Hoehn, K. (2009). *Human Anatomy & Physiology, 8th Edition* (p. 1114). Benjamin Cummings. Retrieved from http://www.amazon.com/Human-Anatomy-Physiology-8th-Edition/dp/0805395695

[31] Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., & Lancaster, J. (1995). A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). *NeuroImage*, *2*(2), 89–101. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9343592

[32] Mori, S., Crain, B. J., Chacko, V. P., & van Zijl, P. C. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of neurology*, *45*(2), 265–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9989633

[33] Moseley, M. E., Kucharczyk, J., Asgari, H. S., & Norman, D. (1991). Anisotropy in diffusion-weighted MRI. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *19*(2), 321–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1652674

[34] Mukherjee, P., Berman, J. I., Chung, S. W., Hess, C. P., & Henry, R. G. (2008). Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. *AJNR. American journal of neuroradiology*, *29*(4), 632–41. doi:10.3174/ajnr.A1051

[35] Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series)* (p. 1104). The MIT Press. Retrieved from http://www.amazon.com/Machine-Learning-Probabilistic-Perspective-Computation/dp/0262018020

[36] Naggara, O., Oppenheim, C., Rieu, D., Raoux, N., Rodrigo, S., Dalla Barba, G., & Meder, J.-F. (2006). Diffusion tensor imaging in early Alzheimer's disease. *Psychiatry research*, *146*(3), 243–9. doi:10.1016/j.pscychresns.2006.01.005

[37] Neal, R. M. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, *9*(2), 249–265. doi:10.2307/1390653

[38] Pele, O., & Werman, M. (2009). Fast and robust Earth Mover's Distances. *2009 IEEE 12th International Conference on Computer Vision* (pp. 460–467). IEEE. doi:10.1109/ICCV.2009.5459199

[39] Pitman, J. (2002). Combinatorial Stochastic Processes. University of California at Berkeley, Berkeley, California: Statistics Department, University of California, Berkeley. Retrieved from http://stat-reports.lib.berkeley.edu/accessPages/621.html

[40] ROSENBERG, D. R., MIRZA, Y., RUSSELL, A., TANG, J., SMITH, J. M., BANERJEE, S. P., BHANDARI, R., et al. (2004). Reduced Anterior Cingulate Glutamatergic Concentrations in Childhood {OCD} and Major Depression Versus Healthy Controls. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(9), 1146–1153. doi:http://dx.doi.org/10.1097/01.chi.0000132812.44664.2d

[41] Radua, J., Phillips, M. L., Russell, T., Lawrence, N., Marshall, N., Kalidindi, S., El-Hage, W., et al. (2010). Neural response to specific components of fearful faces in healthy and schizophrenic adults. *NeuroImage*, *49*(1), 939–46. doi:10.1016/j.neuroimage.2009.08.030

[42] Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision*, *40*(2), 99–121. doi:10.1023/A:1026543900054

[43] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)* (p. 176). Chapman and Hall/CRC. Retrieved from http://www.amazon.com/Estimation-Statistics-Analysis-Monographs-Probability/dp/0412246201

[44] Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., et al. (2004). Advances in functional and structural MR image

analysis and implementation as FSL. *NeuroImage*, *23 Suppl 1*, S208–19. doi:10.1016/j.neuroimage.2004.07.051

[45] Socher, R., & Manning, C. D. (2011). Spectral Chinese Restaurant Processes : Nonparametric Clustering Based on Similarities. *Proceedings of the 14th International Con- ference on Artificial Intelligence and Statistics (AISTATS) 2011* (Vol. 15).

[46] Sondergaard, T. (2011). Data Assimilation with Gaussian Mixture Models using the Dynamically Orthogonal Field Equations by, (2008).

[47] Stejskal, E. O., & Tanner, J. E. (1965). Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics*, *42*(1), 288. doi:10.1063/1.1695690

[48] Sudderth, E. B. (2006). *Graphical Models for Visual Object Recognition and Tracking by*. Massachusetts Institute of Technology.

[49] *The EM Algorithm and Related Statistical Models (Statistics: A Series of Textbooks and Monographs)*. (2003). (p. 250). CRC Press. Retrieved from http://www.amazon.com/Algorithm-Related-Statistical-Models-Statistics/dp/0824747011

[50] Tournier, J.-D., Calamante, F., Gadian, D. G., & Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, *23*(3), 1176–85. doi:10.1016/j.neuroimage.2004.07.037

[51] Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., & Wedeen, V. J. (2002). High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *48*(4), 577–82. doi:10.1002/mrm.10268

[52] Tungaraza, R. F., Howbert, J. J., Mehta, S. H., Haynor, D. R., Shapiro, L. G., & Grabowski, T. J. (2012). IDENTIFYING THE STRUCTURAL ARCHITECTURE OF THE HUMAN INFERIOR PARIETAL LOBULE USING DIFFUSION MRI Departments of 1 Radiology , 2 Psychology , 3 Computer Science and Engineering , and 4 the Integrated Brain Imaging Center , University of Washington , and 5, 506–509.

[53] Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm, *20*(1), 45–57.

[54] Zilles, K., & Amunts, K. (2010). Centenary of Brodmann's map--conception and fate. Nature reviews. Neuroscience, 11(2), 139–45. doi:10.1038/nrn2776