A Multi-resolution Framework for
Statistical Analysis of Neuroimaging Data

by

Won Hwa Kim

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

Doctor of Philosophy
(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN–MADISON

2017

Date of final oral examination: 08/11/2017

The dissertation is approved by the following members of the Final Oral Committee:

Vikas Singh, Associate Professor, Biostatistics and Medical Informatics
Charles R. Dyer, Professor, Computer Sciences
Sterling C. Johnson, Professor, School of Medicine
Barbara B. Bendlin, Associate Professor, School of Medicine
Mohit Gupta, Assistant Professor, Computer Sciences

To my family.

ACKNOWLEDGMENTS

It has been 6 years since I came to Wisconsin on Aug. 11th in 2011. I still remember the moment when I first set my foot in Madison. There were good times and tough times throughout the 6 years, and I couldn't have finished my Ph.D without so much help from so many others. I will spend the next few pages to describe a little bit about my journey for my Ph.D and send out my appreciation.

My trip for Ph.D started a long time ago when I played a board game back in 1988 that I got from one of my father's colleague. The game was about pursuing degrees in universities; I still have no idea who invented such kind of a game. I played it only once where I ended up with a Ph.D, and I thought I would get a real one myself someday. At that time, I had no clue what it meant to get a Ph.D although my father was a professor in Mathematics and had a Ph.D at that time. Later, after spending two years in the U.S. (one year in 1995 and another in 2001) when my father was a visiting professor at universities in Utah, I decided why not get it in the U.S. instead of my home country, South Korea. After 10 years from 2001, I finally got an admission for a graduate school from Computer Sciences at the University of Wisconsin - Madison.

The graduate program was very tough at first, like always, but everything magically worked out after I met my advisor, Prof. Vikas Singh. I was very fortunate to meet Vikas, we ran into each other in AI seminar in September 2011. I've worked with and benefited from him for almost the entire 6 year period for my Ph.D. I would call him my "academic father". I got so much support from him to pursue my research; I did not know what a real "research" was before I met him. I had desire for research when I first met him, and I really appreciate him for guiding me to put all the desire and passion to the right direction (which I think is the most fundamental role as an advisor). Thanks to Vikas, I learned how to come up with creative ideas through discussions and how they should be packaged and written as an academic article to be published to others. He also cared a lot for my family. He gave a warm welcome when my wife, Yeojin, came to Madison, and he was so happy for me when my son,

Jinwoong (Jeremy) was born. I will also never forget the dinner he and Lopa (Vikas' wife) prepared when my parents were visiting Madison. For all the good and bad things happened in my family, he would enjoy and worry as if they were his own matters.

I also needed support from many others other than just Vikas to successfully complete my Ph.D. Actually, my first project with Vikas on action detection completely failed and that was when I turned to looking into neuroimaging researches for Alzheimer's disease (AD). Prof. Sterling C. Johnson, leading the Wisconsin Alzheimer's Disease Research Center (W-ADRC), was kind enough to provide all the support (both by himself and through his lab) for me to learn about AD and neuroimaging. I appreciate him for valuing my work and being very patient through all the meetings we had together. Along with Sterling, Prof. Barbara B. Bendlin was always there to help me work on interesting topics. Most of my works on brain connectivities were with her, she is a magician that turns crazy ideas into very useful research topics. She has been always encouraging, and I would get her positive energy whenever I talked to her. Prof. Charles R. Dyer was like a very nice neighbor next door although he was already a big guy in Computer Vision. I learned so much just talking to him, and I would like to become a mentor like him someday. I hope he enjoys his retirement after so much that he has done for the school and community. I'd like to send my thanks to Prof. Mohit Gupta, who recently joined the department. I remember the lunch with him during his interview here at the University of Wisconsin - Madison, and I was so glad that he made it here. Although we could not spend much time together (since he is new), he gave me so many valuable pieces of advice for my career, and I look forward to meeting him at conferences in the future. Last but not least, I cannot help stress how much impact Prof. Moo K. Chung had on my research. Many of my initial researches were inspired from his previous work and followed the roads that he had already paved. I bugged him so much for his comments and to get his hands for collaboration. A good portion of my knowledge on statistical brain image analysis as well as harmonic analysis comes from him. Coming from the same home country, South Korea, although we spoke English to each other, I felt comfort whenever I talked to him. I was able to complete my first

paper (NIPS 2012) thanks to Moo, which I consider as my best paper during my Ph.D program. Also, the MICCAI 2013 and NeuroImage 2015 papers were inspired when I was taking his course which set a solid corner stone for this thesis (in Chapter 4).

Not only just my Ph.D committee members and mentors above but also my colleagues in our group gave me huge support. Chris Hinrich, who basically lived in the office taught me all about the beer. We sometime went out for drinks on Friday nights when we were the only ones in the office. He was like a big brother here in Madison. Maxwell Collins was a computer guru who helped me to set up lots of experiments when I was not familiar with linux system. He also tipped a lot in restaurants, I can't imagine how much he would give out tips now that he is working in Google. Deepti Pachauri and I were office-mates and we talked about lots of stuffs as if we were sisters. I am so glad that she got married after all. Jia Xu was a solver for almost all problems. He would always come up with good choices when I was not able to make decisions. He is also the only member in our group who I personally met in South Korea when he was visiting Seoul where we hung out together. Vamsi Ithapu and I joined the group in the same semester. We were like soldiers trying to survive through so many difficult courses together. He knows so much about everything and I enjoyed listening to the discussions that Chris and Vamsi had over the beers. Hyunwoo Kim and I are from the same country and we both came to Madison in the same year, 2011. He is one of those crazy Koreans who is particularly fun to have drinks, and we had so much fun at his place in the first year when he and Dongyoung Cho lived in the same apartment. Actually, these guys started a fire in my apartment on my birthday — I should not forget this. Sathya Ravi is an interesting guy, he is very good at both Math and Dota. I don't know what I am going to do without him when I run into problems in optimization in my future research. Seong Jae Hwang was a very good coffee-mate. We would go for coffee every day no matter what, I will miss the coffee breaks I had with him. He is also the one that I can trust and call for help at any time; he even helped me throw away my heavy sofa at my home today. Ronak Mehta was the most gentle and the nicest person. He was the only person that I would say "normal" in our group.

Yunyang Xiong knew more about K-pop than I did. I am sorry that I could not throw a Korean party that you wanted which I still do not know what it is. I hope Ligang Zheng, a visiting professor at our group, enjoys his rest of time in Madison, and Mona Jalal / Tuan Dinh (also my next door neighbor) be successful through their graduate school. And I should not forget Nagesh Adluru; I put him in the last since his last name starts with 'A'. We worked on many papers together and cared each other like brothers. Now that he has decided to settle in Madison, I hope he enjoys his career and life here although I won't be there for him.

Next, I must give my very special thanks to my wife, Yeojin Lee. I got my admission to UW - Madison after 2 months I started dating her. I still don't know what she was into, but when I asked her that I might have to go out side of country for graduate school, she said that she would come with me. At that time I made three promises, 1) I will get my Ph.D in no more than 5 years, 2) I will become a professor in S. Korea, and 3) We will visit Hawaii after my Ph.D. I am so sorry that I was not able to keep them. It took 6 years and we are staying in the U.S. (for now), but at least we visited Hawaii in the final year (to attend CVPR 2017). All the progress I've made so far and I will make, I owe them to her. I love you.

During my Ph.D, not only I ended up with a dissertation but I was also productive in another way: I got my son Jeremy Jinwoong Kim in 2013. I had to give up my trip to CVPR 2013 where I had my first paper in Computer Vision venue. I was extremely looking forward to it but my presentation schedule (Jun. 25th, 2013) was on the same day that Jinwoong was expecting to see the world for the first time. When we first met, I felt like who cares about CVPR? I actually have almost empty publication list in 2014 since I did nothing but watching him all the time. He is the most valuable project that I worked on (with Yeojin) in Madison, and I will put my full support if he decides to purse a Ph.D someday.

My mother (Soon Hee Choi) and father (Anhyun Kim), coming at last since they are very important to acknowledge, are at the root of who I am now. My father put the seed in me to grow up to be a Ph.D and my mother watered the seed with so much support. Both my parents being teachers, they never

hesitated supporting me from learning and pursuing my study through so many schools until today. Frankly, I never thought I would be in school this long until my mid-thirty and my brother, Won Yong Kim, once told me, "You grew up and became a student." Here with submission of this thesis, I put a period to conclude my student life and open up the next chapter in my life.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Statistical analysis of brain images/image-derived measures plays a central role in discovering associations between different brain regions and covariates. When this framework is used to analyze neurodegenerative diseases such as Alzheimer's disease (AD), the major goal is to identify, ideally early on, which of the brain regions show abnormal variations due to the disease, so that we can provide intervention and treatment to slow down the progression of the disease. Unfortunately, there are several factors that make this statistical analysis problematic. In most brain imaging studies, the sample size is limited (typically up to only a few hundreds) due to the high cost of scans and difficulties in recruiting participants depending on diseases or risk factors. In many cases, thus, it may not be sufficient to robustly achieve statistically meaningful outcome especially when the effect size is small in the preclinical stages. Moreover, there are many nuisance factors that affects the analyses. To deal with the challenges above, in this thesis, we propose novel multi-resolution frameworks which we will develop and experimentally evaluate on a variety of neuroimaging data. These frameworks make use of recent work from harmonic analysis literature which implement "wavelet transform" in non-Euclidean spaces, so that we can adopt the multi-resolution scheme not only for imaging data in the Euclidean space (i.e., $\mathbb{R}^n$) but for image derived measures represented in non-Euclidean spaces such as cortical thickness on brain meshes and brain connectivity. We describe the algorithmic development and how such methods can help evaluate novel scientific hypothesis. For each framework, we demonstrate extensive experimental results to show that the frameworks improve statistical outcome over traditional approaches and can be easily adopted for real data analyses.

1 INTRODUCTION

1.1 Analysis of Neuroimaging Data

Brain imaging plays a major role in clinical practice and scientific research in neuroscience. A broad spectrum of neuroscience studies heavily rely on statistical analysis of the brain images to focus on understanding the process of aging (Jagust et al., 2006; Wingfield and Grossman, 2006; Dennis and Cabeza, 2008), the effects of trauma (Williams et al., 2006; Hull, 2002) and the manifestation of neurological disorders (Savitz and Drevets, 2009; Castellanos et al., 2002). In these analyses, one important question is to identify which specific brain regions are affected as a function of disease and/or to find associations of regional measurements in the image with future cognitive decline.

There are various types of imaging modalities used in scientific and clinical research, and each of them characterize different aspects of the brain. Magnetic resonance image (MRI) is one of the most commonly used imaging modalities in neuroscience which is based on the magnetization properties of atomic nuclei in the brain (Haacke et al., 1999). An external magnetic field is applied to the brain to align the protons that are randomly oriented within the water nuclei of the tissue, and by measuring the relaxation time taken for the nuclei to return to their resting alignment, we obtain T1 and T2 weighted images as shown in the left of Fig. 1.1 (Johnson, 2008; Haacke et al., 1999; Ogawa et al., 1990). MR images provide detailed information on grey matter containing dendrites and axon terminals of neurons and white matter made of axons connecting different parts of grey matter. It is typically used to analyze structural pathologies of brain diseases (Courchesne et al., 2001; Watkins et al., 2002; Sparks et al., 2002; Jack et al., 2004).

Also, advances in diffusion tensor imaging (DTI), which is a form of MR Imaging, paved the way for many important *in vivo* investigations of white matter microstructure (Uhlhaas and Singer, 2006; Crossley et al., 2014; Konrad and Eickhoff, 2010). By applying diffusion encoding gradients to the brain, one measures the orientation of water molecules in the brain tissue as a "diffusion tensor" describing the diffusion direction at each voxel of the image. It is a non-

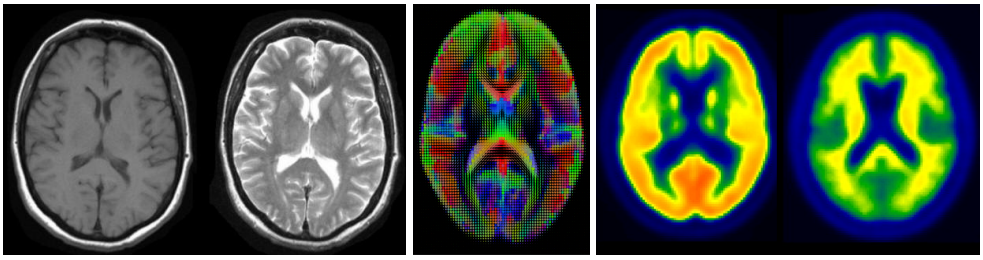


Figure 1.1: Examples of neuroimages. Left: T1-weighted and T2-weighted MRI capturing grey and white matters, Middle: DTI capturing the orientation of water molecules, Right: FDG-PET and PiB-PET images capturing metabolism in the brain.

invasive imaging method to characterize the microstructural organization of tissue, and offers information on the tissue microstructure (Jones and Leemans, 2011). An example of a DTI scan is shown in the middle of Fig. 1.1.

While the aforementioned imaging modalities provide structural information in the brain, positron emission tomography (PET) scans characterize metabolic changes at the cellular level in the brain tissue. In many cases, a disease process begins with functional changes at the cellular level in the very early stages of the disease and PET images can capture these changes before the brain exhibits structural variations shown in the MRI. To acquire a PET scan, a radiotracer is injected into the bloodstream which travels to organs and emits gamma rays indicating tissue metabolic activity. By observing the metabolic rate in the brain using different types of the radiotracers, we can capture various pathologies of diseases. For example, Pittsburgh compound B (PiB) PET captures uptake of tissue radioactivity concentration (Ikonomovic et al., 2008; Rinne et al., 2010) and Fludeoxyglucose (FDG) measures the uptake of glucose in brain tissue (Chao et al., 2001; Padma et al., 2003), and representative images of a FDG PET scan and a PiB PET scan are shown in the right of Fig. 1.1.

Once the neuroimaging data are collected, one may apply statistical analysis methods directly on the acquired images, or one can pre-process these brain images to obtain image-derived measures that characterize other types of information from the brain. For instance, running the Freesurfer algorithm (Fischl, 2012) on a MRI scan yields inner and outer cortical surfaces as shown

in Fig. 1.2. By measuring the distance between the two surfaces, we can obtain cortical thickness measures on a brain surface which is a relevant feature in many brain disorder analyses. The existing literature tie this measure to brain growth (O'Donnell et al., 2005; Shaw et al., 2006a; Sowell et al., 2004; Lemaitre et al., 2012), autism (Chung et al., 2005), attention-deficit (Shaw et al., 2006b), genetic influences (Panizzon et al., 2009), amusia (Hyde et al., 2007), osteoporosis (Hodsman et al., 2000), and even gender (Sowell et al., 2007).

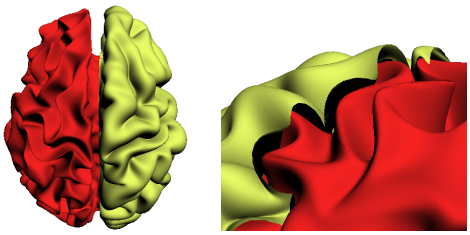


Figure 1.2: Illustration of cortical thickness. The inner cortical surface (red) is covered by the outer cortical surface (yellow), and the cortical thickness is measured by the distance between the outer and the inner cortical surfaces.

The recent development of tractography with DTI data offers a characterization of the human connectome (the brain's wiring diagram) to better understand the *structural* aspects of brain connectivity (Le Bihan et al., 2001). DTI measures macroscopic axonal organization in nervous system tissue, which corresponds to complex neural fiber network in human brains. Performing tractography on a DTI image yields structural brain connectivity at the individual level, and various parameters are defined along the neuron fiber bundles that correspond to the connection between different regions of interests (ROIs) measuring the strength of the connection (e.g., number of fibers, fractional anisotropy (FA) and mean diffusivity (MD)) (Basser et al., 2000). These measures have been used to analyze traumatic brain injury (TBI) (Niogi et al., 2008; Wozniak et al., 2007), Parkinson's disease (Vaillancourt et al., 2009) and perform brain analysis of infants and children (Glenn et al., 2003).

Regardless of the types of the data, in an ideal case, a statistical analysis method is applied at each data point location (e.g., voxel in a 3D volume image) on the data collected from a population of subjects. The objective of such an analysis may be to identify variations due to a certain variable of interest (e.g., a disease, a risk factor and a treatment) that are statistically significant. Let us first consider a simple case where we observe a single measurement (e.g., age or a summary measure of the whole brain) per subject. If the cohort can

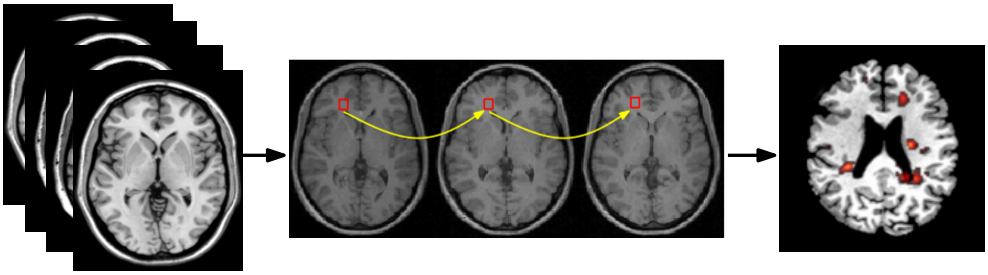


Figure 1.3: Voxel-wise brain image analysis. Brain scans from different participants (left) are first registered to a template image (middle) to obtain voxel-to-voxel correspondence, then statistical test is performed at each voxel to obtain resultant p-value map showing clusters are shown on the template brain (right).

be stratified into two (or more) groups and the groups are matched by other confounding variables that may affect the final outcome, a group analysis (e.g., diseased versus healthy controls) is performed on the stratified data using a hypothesis test (e.g., two sample t-test). For the hypothesis test, we basically assume that there is no difference between the two groups and the statistical test examines whether the assumption should be accepted or rejected. A p-value is returned as a result of the hypothesis test, and one can conclude that there is a significant group difference when the p-value is sufficiently low (e.g., $\leqslant 0.05$). If the groups are not matched by other covariates or we are interested in a variable that is continuous and does not indicate the group information (e.g., age and cognitive score), we can use regression models (e.g., general linear model, multi-variate general linear model and mixed effects model) to find out the effect from the variable of interest.

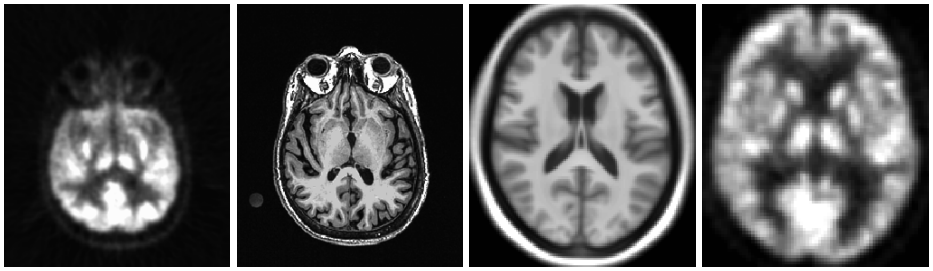
Performing such statistical analysis on neuroimages is complicated due to many reasons. First, human brains may have different sizes and shapes. Since the same voxel location in different images may correspond to different parts of the brain, the images from each individual cannot be compared directly. Therefore, these images must be registered on a common coordinate space to obtain "voxel-to-voxel correspondence". The registration process enables us to perform meaningful statistical analysis at each voxel (or data point) with a variable of interest to obtain p-values for every voxel. But notice that we have to perform the statistical test as many times as the number of data

points (i.e., voxels). In this case, we face the so-called multiple testing problem (Dunnett, 1955; Hsu, 1996; Keppel and Wickens, 2004), i.e., an error from each test accumulates as the number of tests increases which eventually leads to many false-positives. The multiple comparisons issue can be corrected using a threshold on the p-values that takes the number of total tests into account. When the correction is done and the resultant p-values surviving the multiple comparisons correction are projected on the template space, they may form "clusters" of voxels that indicate regions in the brain that are affected by the variable of interest with statistical evidence. Such a pipeline is demonstrated in Fig. 1.3.

1.2 Challenges in the Analysis of Neuroimaging Data

Despite advances in image acquisition techniques and sophisticated analysis methods to better understand various aspects of the human brain, there are several challenges that the standard neuroimaging/neuroscience studies routinely face. The most fundamental challenge is to recruit a sufficient number of subjects. Even if there is a clear disease specific variation in the brain, it is difficult to characterize the existing variations using a statistical method with insufficient data (i.e., small sample-size). Therefore, the cohort size must be large enough to ensure that we can reliably identify the variation. However, in many neuroimaging studies, this may not be feasible due to cost and/or the specific scope of the clinical question of interest (e.g., demographic requirements, genetic profile) which restricts the number of participants, typically, to no more than a few hundred. In this small sample-size regime, the analysis becomes particularly problematic when the effect size of interest is weak or subtle. Since we are progressively moving towards analyzing early symptoms of disease and often, even in the pre-clinical stages, this is a critical issue that needs to be addressed. Therefore, it is imperative that the analysis method is *sensitive* enough to detect even the subtle signal variations in the brain.

Also, accurate *registration* of the brain images is necessary but very difficult to achieve in reality. Correct registration is critical in voxel-wise statistical image analysis, since it provides voxel-to-voxel correspondence among different



(a) (b) (c) (d)

Figure 1.4: Example of multi-modal image registration for statistical analysis. a) FDG-PET scan, b) T1-weighted image, c) Dartel T1 template, d) Registered FDG-PET scan from a). Direct registration of a) to c) is problematic. Therefore, a) is first co-registered to b), and the warping of T1-weighted image to the template (from b) to c)) is computed. Then the warping is applied to a) to obtain a registered PET image as in d).

images in the dataset to perform meaningful voxel-wise analysis (Jenkinson and Smith, 2001; Maes et al., 1997; Pluim et al., 2003; Wells III et al., 1996). However, even the state-of-the-art (linear or non-linear) registration techniques sometime fail to register different images across a dataset when the resolution is poor and it is often problematic to register images between different modalities (Viola and Wells III, 1997; Ashburner and Friston, 1997; Roche et al., 1998). For example, as demonstrated in Fig. 1.4, co-registering two PET images requires additional steps compared to registering MRI scans due to its low resolution. In order to co-register two PET images, the images are first registered to the T1-weighted images of the same subject and then the T1-registered images are warped to each other for registration. Such a process is complicated and may involve registration error, and thereby would result in significant decrease in statistical power. In some cases, such as infant/adolescents brain analysis, a specific template for the population may not even be available leading to sub-optimal results.

There are often issues arising from *systematic variations* in the imaging data as well, which are caused by various factors in the data acquisition process (Colcombe et al., 2003). For example, in a longitudinal image study, the same participant goes through multiple scans at different time points. When there are systematic differences between the scanning process at different times such as changes in scanner, scanner parameter or imaging protocol, such factors will cause systematic variations in the acquired images. Such systematic

variations also cause problems in multi-site studies where the data are collected in different locations (Jovicich et al., 2006) or datasets from multiple locations are concatenated in an effort to increase the sample size (Zhou et al., 2016, 2017). In this regime, it is quite challenging is to capture only the disease specific phenomena under study and minimize the effect of confounding factors.

In order to identify the full relationships among different variables (i.e., ROIs and covariates) a graphical model selection method is often adopted which estimates a "sparse" precision matrix (i.e., inverse covariance matrix) (Banerjee et al., 2006; Honorio and Samaras, 2010) whose non-zero elements describe the conditional dependencies among the variables. Unfortunately, when identifying these associations between brain regions and covariates, we often face a problem caused by *latent variables* that affects our observation (i.e., measurements / data) yet hidden. In many cases, the measurements (image intensity or image-derived measurements) obtained from the ROIs may be dependent on unknown factors (i.e., latent variables) and prevent one from identifying the true relationships between the covariates by making the precision matrix "dense". This is especially true in many real datasets which involve a large list of latent variables affect the observed variables but remain unknown. Accounting for such latent variables is not clearly addressed by standard precision matrix estimation, and is typically approached using specialized optimization techniques with many constraints (e.g., number of latent variables) (Chandrasekaran et al., 2012; Marlin and Murphy, 2009).

Last but not the least, many recent brain imaging datasets contain data that live in non-Euclidean spaces (e.g., a graph that consists of vertices and edges). For example, cortical thickness measures (see Fig. 1.2) on brain surfaces can be considered as signals defined on a graph where the signal is the thickness measure and the graph is the brain surface mesh. Also, the tractography derived brain connectivity measures from DTI can be naturally represented as a graph where the

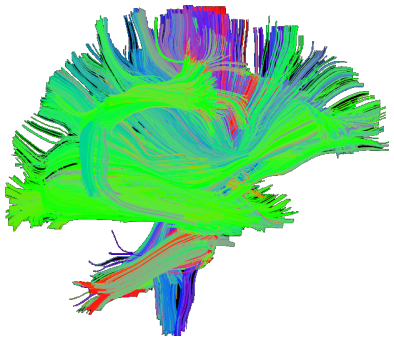


Figure 1.5: An example of neural fiber bundles derived using tractography on DTI.

ROIs correspond to vertices and the edges are defined by the strength measures of the connections from neural fiber bundles. In such settings, where the domain of the data is non-Euclidean, it is problematic to perform traditional image processing techniques to enhance the sensitivity of the method since the domain has an arbitrary structure. For instance, in the Euclidean setting (i.e., uniformly sampled lattice), Gaussian smoothing is performed on each image by kernel convolution to increase signal-to-noise ratio in an image and increase the statistical power. However, if the signal is defined in a non-Euclidean space, it is difficult to define such filtering operations due to the ambiguity of shapes of basis functions to transform the signal into the frequency space.

1.3 Image Analysis in Alzheimer's Disease

Throughout this thesis, we will focus on the analysis of AD using various neuroimaging modalities. AD is one of the most common progressive neurodegenerative diseases characterized by severe loss of memory, decline in cognitive function, changes in mood, inability to find orientation and so on (McKhann et al., 2011) . The disease progresses through multiple stages and its pathology can accumulate before development of clinically relevant cognitive impairment (Perez-Nievas et al., 2013; Chételat et al., 2013). Before being diagnosed as AD, patients go through the mild cognitive impairment (MCI) stage and annually 10% to 15% of the MCI patients convert to AD (Okello et al., 2009; Risacher et al., 2009). Over 24 million people in the world are affected by AD, and unfortunately, there is still no cure for the disease (Leow et al., 2009). It is therefore critical to detect the disease in the early stages and slow down its progression by intervention.

AD poses a challenge given that the cognitive changes that define the disease do not manifest until significant brain pathology has accumulated, and often these cognitive changes are not highly correlated with certain pathological features of the disease. A well-known AD related change in the brain is severe thinning in cortical thickness on several different locations in the brain (Pachauri et al., 2011; Dickerson and Wolk, 2012; Thompson et al., 2011). Given that brain function and pathology manifest strongly as changes in the cortical

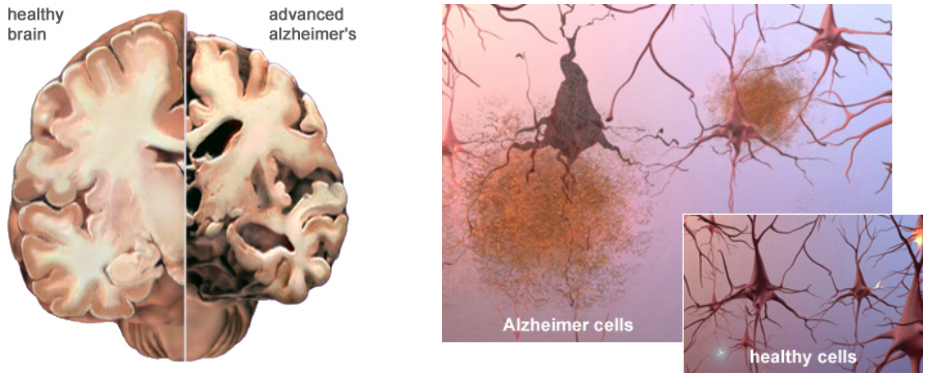


Figure 1.6: Examples of variations due to AD (figures captured from Alzheimer's Associations International Conference (AAIC)). Left: Structural changes (i.e., cell loss) in the brain due to AD, Right: Amyloid plaques (i.e., abnormal clusters of protein fragments) and tangles due to AD.

thickness, the analysis of such data (to find group level differences in clinically disparate populations) plays a central role in structural neuroimaging studies.

Also, post-mortem diagnosis of the disease is determined by regional density of beta-amyloid plaques and neurofibrillary tangles (as shown in the right in Fig. 1.6) with amyloid accumulation (Jack et al., 2013; Montine et al., 2012). It is therefore meaningful to characterize the disease by the changes in neural connectivity captured by different measures, taking into account the neural networks that comprise several affected regions. AD is well-known to be tied to functional connectivity changes as well (Wang et al., 2007; Damoiseaux et al., 2012; Supekar et al., 2008), and therefore changes in brain connectivities can explain behavioral symptoms in AD and may predict conversion to AD (Filippi and Agosta, 2011; Li et al., 2002; Shao et al., 2012).

1.4 Multi-resolution Analysis of Neuroimaging Data

In order to tackle the challenges addressed above arising from the analysis of AD using neuroimaging data, the key concept underlying the methods that we propose is a "multi-resolution" framework in non-Euclidean spaces. In the traditional Euclidean setting, the representation of an image or a signal at different resolutions (as shown in Fig. 1.7) was used to obtain invariance to

Figure 1.7: An example of multi-resolution of an image from Jeremy (my son). Top: Gaussian Pyramid (i.e., high to low resolutions).

scale, which is one of the the most fundamental concepts in computer vision. Its applications span interest point detection (Lowe, 2004; Zhang et al., 2006b; Donahue et al., 2014; Bay et al., 2006), denoising/filtering (Wink and Roerdink, 2004; Sardy et al., 2001; Chang et al., 2000) and compression (Chang et al., 2000; Lu et al., 2000), and is often studied as scale space theory (Lindeberg, 1994). An analog of this concept from the signal processing perspective is "wavelet transform". The wavelet transform uses a *localized (i.e., translated and has finite duration)* oscillating function at multiple *scales*, i.e., mother wavelets, as the basis instead of the sine basis with infinite duration in the Fourier transform. It captures "local context" information (i.e., relationship among neighboring pixels in an image) and overcomes the key limitation of Fourier series in failing to capture sharp changes in a function (i.e., Gibbs phenomena due to infinite support) via the localization property.

Unfortunately, the conventional formulation of wavelet transform has been traditionally studied only in the Euclidean space (e.g., a regular lattice). This is because the construction of the wavelet basis in the non-Euclidean setting is not straight forward due to the ambiguity of scales and translation, which are the key properties of the mother wavelet. Recent literature in harmonic analysis provides an interesting development by defining the wavelet basis as band-pass filters in a dual space (i.e., an analog of the frequency space) where the notion of the scale is easier to define (Hammond et al., 2011; Coifman and Maggioni, 2006). To do this, they first define orthogonal transformations for graphs which is an analog of traditional Fourier transform. When the band-pass filtering operation is implemented in the original space by a inverse transformation from the dual space, it constructs a wavelet basis that satisfies

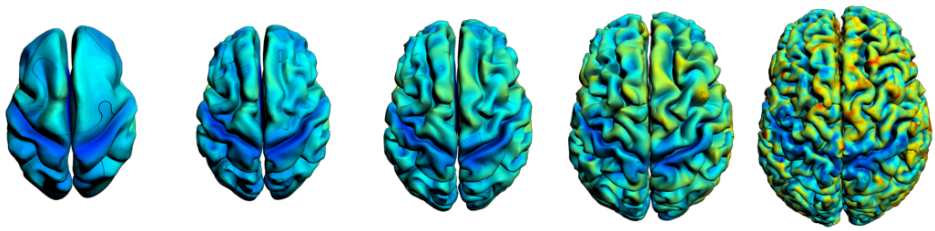


Figure 1.8: An example of multi-resolution in non-Euclidean space. Cortical mesh and cortical thickness signals on the vertices of the mesh are shown in multiple resolutions.

the key properties of traditional mother wavelets. Once the wavelet transform is defined for the signals that are defined on graph vertices, now one can think of this wavelet transform as viewing each vertex of a graph zoomed in-and-out at various resolutions based on the connections among neighboring vertices.

1.5 Aims of the Thesis

Throughout this thesis, we propose various novel multi-resolution frameworks that provide solutions to fundamental issues that routinely arise in neuroimaging data analysis by solving the corresponding technical problems. These frameworks explore the underlying structure of the data to offer desirable properties and benefit their applications in real data analysis. The technical core driving the frameworks is based on the wavelet transform on graphs. Below is a brief description of the frameworks:

- **(i)** A sensitive method for analyzing cortical thickness signals on brain surfaces for AD; the method defines multi-scale descriptor at each vertex of a cortical mesh using wavelet transform on graphs, which benefit the downstream the statistical analysis (Kim et al., 2012, 2014, 2015c).
- **(ii)** A sensitive method for analyzing signals on brain network connection (i.e., edges) for AD and preclinical AD; the method utilizes the dual representation of a graph and the wavelet transform on graphs to derive a multi-scale descriptor at each connection of a brain network which improves the results from statistical analysis (Kim et al., 2013a, 2015a,c).
- **(iii)** A method for performing statistical analysis for longitudinal images with

systematic variations; instead of direct comparison of two longitudinal images with systematic variations, the method defines operators from each image that is invariant to the systematic variations and uses them as a surrogate for detecting changes between them (Kim et al., 2015a).

(iv) A method for performing statistical analysis of images with local deformations; the method defines an image representation that is invariant to the local deformation error using convolution neural network of a graph and the downstream analysis using the new image representation yields more robust results (Kim et al., 2015b).

(v) A method for identifying relationships between different brain neural pathways and relevant covariates using human connectome project (HCP) data; a latent variable graphical model selection method that identifies relationships between different random variables in the presence of unknown latent variables affecting the observation (Kim et al., 2016b).

The full description of these methods is introduced in the subsequent chapters. Together with the presentation of the methods, we provide extensive experimental results demonstrating improved results in the statistical analysis of neuroimaging data and enabling discovery of new scientific findings in AD and preclinical AD analyses.

2 BACKGROUND

2.1 Continuous Fourier and Wavelet Transforms

Let $f(x)$ be a integrable function (or signal) defined in $\mathbb{R}$ parameterized by x. It is natural to examine the signal in the original space x, however, it merely provides the information of the signal regarding its value at a given location in x. Analyzing the signal in x does not convey the information as to the rate at which the signal is varying. To see why this is important, for example, shifting $f(x)$ just a little bit does not affect its shape but changes the $f(x)$ at every location and makes it a very different function if we analyze it at every point in the domain. In such a case, rather than analyzing the signal in the original space, it can be useful to look at a new representation of the signal in a dual space that describes how the signal is varying.

In traditional signal processing, the Fourier transform is the most well-known technique to analyze how a signal is changing (i.e., exploring its frequency characteristics) by decomposing a signal $f(x)$ into a linear combination of Fourier coefficients and Fourier bases (i.e., $\sin()$) (Bracewell and Bracewell, 1986; Welch, 1967; Harris, 1978). A simple example of the Fourier transform is given in Fig. 2.1, where on the left is the original signal $f(x) = \sin(80x) + \sin(300x)$, in the center is a corrupted version of the $f(x)$ with noise and the right is the signal in the freuqency space (i.e., the dual space). As seen on the right in Fig. 2.1, there are two peaks at 30Hz and 100Hz that tells us that the $\sin()$ at those two frequencies are the major components to constitute

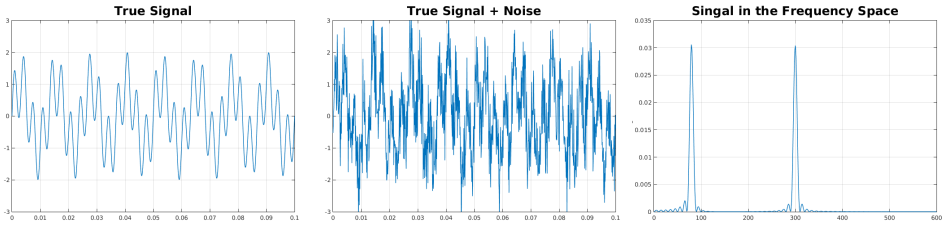


Figure 2.1: An example of Fourier transform. Left: original signal, Middle: corrupted signal (i.e., original signal with noise), Right: corrupted signal in the frequency space.

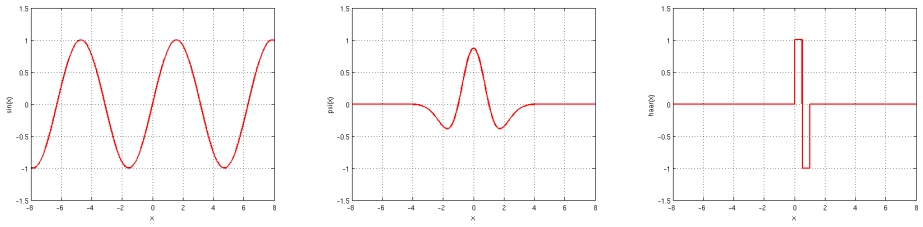


Figure 2.2: Examples of a Fourier basis and mother wavelets. Left: sine basis for Fourier transform, Middle: mexican hat wavelet, Right: haar wavelet. Unlike the sine basis, mother wavelets are localized and have finite support.

the original signal $f(x)$.

Mathematically speaking, the Fourier transform uses $\sin()$ functions as the bases to decompose an original signal $f(x)$. It basically transforms the $f(x)$ to the frequency space denoted by ω as

$$\hat{f}(\omega) = \int f(x) e^{-j\omega x} dx \tag{2.1}$$

resulting in Fourier coefficients $\hat{f}(\omega)$. Its inverse transform reconstructs the original signal as a linear combination of the coefficients and the bases as

$$f(x) = \frac{1}{2\pi} \int \hat{f}(\omega) e^{j\omega x} d\omega. \tag{2.2}$$

The wavelet transform is similar to the Fourier transform in that a signal is decomposed as a linear combination of coefficients and certain basis functions. While the Fourier expansion uses $\sin()$ bases which have infinite support and is localized only in the frequency space, the wavelet expansion instead uses a mother wavelet basis ψ which is localized in *both time and frequency* (Haykin and Van Veen, 2007). Representative examples of Fourier and mother wavelets (i.e., Haar wavelet and Mexican hat wavelet) are shown in Fig. 2.2.

The classical wavelet transform starts by defining a mother wavelet $\psi_{s,a}$ with a scale parameter s and a translation parameter a as

$$\psi_{s,a}(x) = \frac{1}{s} \psi\left(\frac{x-a}{s}\right). \tag{2.3}$$

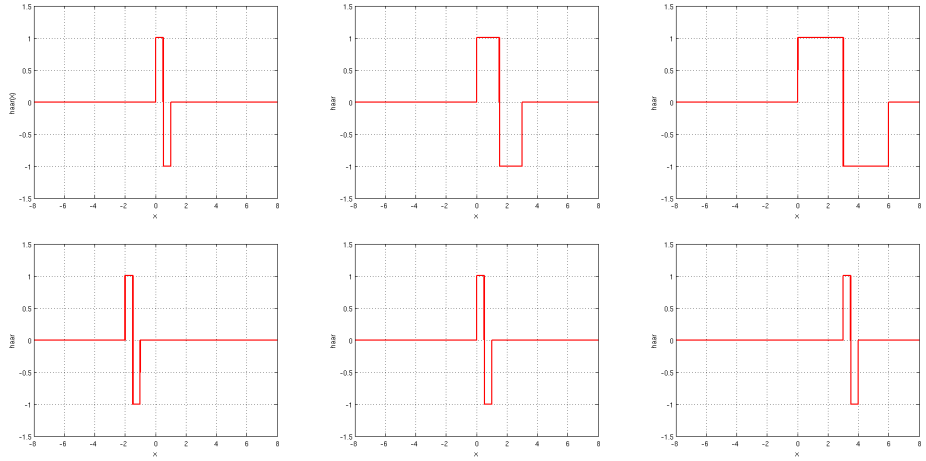


Figure 2.3: Examples of scale and translation properties of haar wavelet. Top row: haar wavelet in different scales (i.e., dilation), Bottom row: haar wavelet in different locations.

Here, the parameter s governs the dilation and the parameter a controls the localization of $\psi_{s,a}$ respectively as shown in Fig. 2.3. Using $\psi_{s,a}$ as the bases, the wavelet transform of a function $f(x)$ is defined as the inner product between the ψ and f, represented as

$$W_f(s,a) = \langle f, \psi_{s,a} \rangle = \frac{1}{s} \int f(x) \psi^*(\frac{x-a}{s}) dx \tag{2.4}$$

where $W_f(s,a)$ is the wavelet coefficient at scale s and at location a, and ψ^* is the complex conjugate of ψ. The inverse wavelet transform can be defined if it satisfies the admissibility condition,

$$C_\psi = \int \frac{|\hat{\psi}(j\omega)|^2}{\omega} d\omega < \infty \tag{2.5}$$

where C_ψ is the admissibility constant and $\hat{\psi}(j\omega) = \int \psi(x) e^{-j\omega x} dx$ is the Fourier transform of the mother wavelet. Once the admissibility condition is satisfied, such a wavelet transform is invertible, and the inverse wavelet transformation reconstructs the original signal $f(x)$ from $W_f(s,a)$ without any

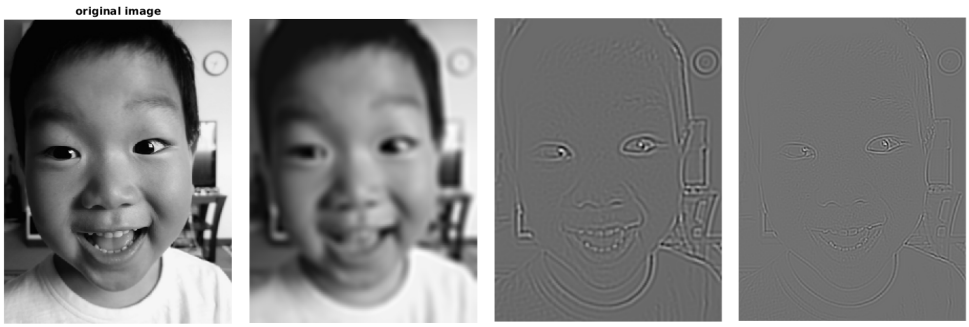


Figure 2.4: An example of wavelet transform of an image. First: original image, Second: wavelet transform using a scaling function, Others: wavelet transform using wavelet bases in different scales.

loss of information as,

$$f(x) = \frac{1}{C_\psi} \iint W_f(s, a)\psi_{s,a}(x)\,da\,ds. \tag{2.6}$$

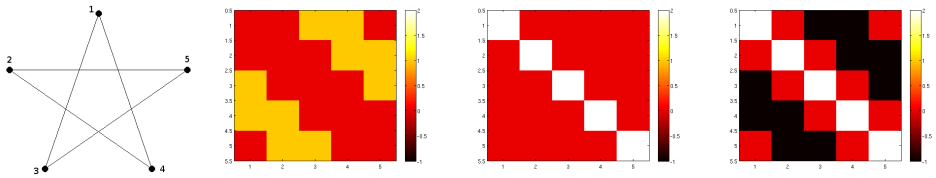
Here, (2.6) is known as *resolution of the identity* and the key expression for multi-resolutional analysis using the wavelet transform (Haykin and Van Veen, 2007; Daubechies, 1990).

Interestingly, the mother wavelets ψ_s at multiple scales behave as band-pass filters corresponding to different bandwidths in the frequency space. When these band-pass filters do not cover the low-frequency components, an additional scaling function ϕ (i.e., father wavelet) that behaves as a low-pass filter is introduced, and a transform with the scaling function ϕ returns a low-pass filtered (i.e., smooth representation) of the original function f. Due to this selective filtering property, wavelets offer a multi-resolution view of the given signal. A classical example of the wavelet transform is demonstrated in Fig. 2.4 with an original image, wavelet trasform using the scaling function and mother wavelets in multiple scales.

2.2 Fourier and Wavelet Transforms in Non-Euclidean Spaces

As previously described in Section 2.1, the Fourier transform and the wavelet transform of a signal in the Euclidean spaces (i.e., $\mathbb{R}^n$) are well defined and have been popular in signal processing for representing the signal in the frequency space to obtain its frequency characteristics (Haykin and Van Veen, 2007; Bracewell and Bracewell, 1986; Takeda et al., 1982; Daubechies, 1990; Antonini et al., 1992). The implementation of these transforms in the Euclidean space is convenient thanks to the *regularity* of the domain. Since the domain is typically represented as a regular lattice, one can easily define the 'shape' of the basis (i.e., $\sin()$ or a mother wavelet). However, in a non-Euclidean setting where the domain is *irregular* (e.g., a graph), the definition of the basis in such a complex domain becomes difficult. For example, on a graph with a set of vertices and edges connecting the vertices, the notions of scale and translation of a mother wavelet $\psi_{s,a}$ are not as easy to conceptualize when the distance between each vertex and the number of connected edges are not uniform. Due to this difficulty, the wavelet transform has not been suitable for the analysis of signals when the domain has an arbitrary structure until recently when (Coifman and Maggioni, 2006; Hammond et al., 2011; Narang et al., 2012) presented a result dealing with wavelet and Fourier transform of graphs (and other non-Euclidean spaces).

Formally, a graph $G = \{V, E\}$ is defined by a vertex set V (where the number of vertices is N) and a edge set E. Such a graph G is generally represented as an adjacency matrix A of size $N \times N$ where each element a_{ij} denote the edge weight between the ith and the jth vertices. Another graph representing matrix is a degree matrix D, which is a diagonal matrix with the ith diagonal as the sum of edge weights connected to the ith vertex. From the adjacency matrix and the degree matrix, a graph Laplacian is defined as $L = D - A$, which is an analog of the Laplacian operator in the continuous space. An example with a star shaped graph and matrices representing the graph (i.e., its adjacency and degree matrices and graph Laplacian) are shown in Fig. 2.5. Here, the graph Laplacian L is self-adjoint and positive semi-definite, therefore the spectrum



(a) (b) (c) (d)

Figure 2.5: An example of Spectral Graph Theory. a) A star shaped graph, b) Adjacency matrix A representing connections between the vertices, c) Degree matrix D, d) Graph Laplacian $L = D - A$.

of L yields pairs of eigenvalues $\lambda_l \geqslant 0$ and the corresponding eigenvectors χ_l where $l = 0, 1, \cdots N-1$. The χ_l are orthonormal eigenvectors that can be used as analogs of $\sin()$ bases in the traditional Fourier transform. Examples of χ_l defined on a sphere shaped mesh are shown in Fig. 2.6, where the color represents the function values. Notice that there are peaks represented in red and blue which correspond to the high and low peaks of traditional $\sin()$ functions, and as the order l increases, the number of peaks increases indicating higher frequency.

Given the general definition of a inner product for two vectors (i.e., for complex vector) **a** and **b** as

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum a_i b_i^*, \tag{2.7}$$

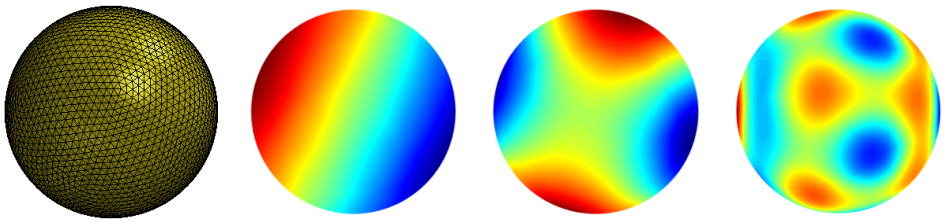
the graph Fourier transform using χ_l as the bases is defined as an inner product between the function $f(n)$ on the graph nodes and a basis vector $\chi_l(n)$ as

$$\hat{f}(l) = \langle f, \chi_l \rangle = \sum_{n=1}^{N} f(n) \chi_l^*(n) \tag{2.8}$$

where $\hat{f}(l)$ is the graph Fourier coefficient and the inverse transform is defined as

$$f(n) = \sum_{l=0}^{N-1} \hat{f}(l) \chi_l(n). \tag{2.9}$$

Notice that (2.8) and (2.9) are similar to (2.1) and (2.2) in traditional Fourier



(a) (b) (c) (d)

Figure 2.6: Examples of graph Fourier bases on a sphere mesh. Observe that as the order of the basis increases, the frequency of fluctuation of the basis changes (i.e., higher order corresponds to higher frequency). a) A 3-D triangular mesh domain (sphere) with 2562 vertices and 5120 faces, b) χ_2, c) χ_9, d) χ_{27}.

transform but use different orthogonal bases in the discrete graph space. This transform offers a convenient means to transform a signal/measurement on graph vertices to a dual domain which is an analog of the frequency domain in traditional Fourier analysis.

Utilizing the graph Fourier transform, the mother wavelet ψ can be constructed by first defining a kernel function $g()$ (i.e., band-pass filter) in the frequency domain and then localizing its operation by a delta function δ in the original graph space via the inverse graph Fourier transform. In other words, a band-pass filter $g()$ is first defined in a graph Fourier space defined by λ_l, and then a ψ is constructed when the band-pass filtering operation by $g()$ is implemented in the original graph space by inverse graph Fourier transform. Since the frequency representation of a δ_n is $\langle \delta_n, \chi_l \rangle = \chi_l^*(n)$ (using (2.7)) in the dual space, the mother wavelet $\psi_{s,n}$ at vertex n at scale s is constructed (i.e., localized) using the inverse Fourier transform (2.9) as

$$\psi_{s,n}(m) = \sum_{l=0}^{N-1} g(s\lambda_l)\chi_l^*(n)\chi_l(m) \tag{2.10}$$

which is a band-pass filtering of a delta function δ_n in the graph Fourier space. Notice that the scale s is defined *inside* $g()$ by the scaling property of Fourier transform (Haykin and Van Veen, 2007) and the eigenvalues λ_l serve as the analogs of frequency. Also, since $L \succeq 0$, $\chi_l^* = \chi_l$ and $\psi_{s,n}(m)$ can be

considered as a symmetric kernel defining a distance between two vertices n and m (i.e., $\psi_{s,n}(m) = \psi_{s,m}(n)$). Examples of $\psi_{s,n}$ (i.e., Mexican hat wavelet) in different scales are shown in Fig. 2.7.

Using $\psi_{s,n}$, the wavelet transform of a function $f(m)$ at scale s can be easily defined as an inner product using the mother wavelets $\psi_{s,n}$ as

$$
\begin{aligned}
W_f(s,n) &= \langle f, \psi_{s,n} \rangle && (2.11)\\
&= \sum_{m=1}^{N} f(m)\psi_{s,n}(m) && (2.12)\\
&= \sum_{m=1}^{N} f(m)\psi_{s,m}(n) \quad \text{(using the symmetry of } \psi) && (2.13)\\
&= \sum_{m=1}^{N} f(m) \sum_{l=0}^{N-1} g(s\lambda_l)\chi_l^*(m)\chi_l(n) && (2.14)\\
&= \sum_{l=0}^{N-1} g(s\lambda_l)[\sum_{m=1}^{N} f(m)\chi_l^*(m)]\chi_l(n) && (2.15)\\
&= \sum_{l=0}^{N-1} g(s\lambda_l)\hat{f}(l)\chi_l(n) && (2.16)
\end{aligned}
$$

resulting in wavelet coefficients $W_f(s,n)$ just like in (2.4). Such a transform offers a multi-resolution view of signals on graphs, and the multi-resolution property can be easily captured by a single parameter s in the kernel function $g()$.

Note that this expression (2.16) corresponds to the continuous wavelet transformation as shown in (2.4) with an integral of a set of coefficients and given wavelet bases. If the kernel $g()$ satisfies the admissibility condition

$$C_g = \int_0^\infty \frac{g^2(x)}{x} dx \leqslant \infty \tag{2.17}$$

and $g(0) = 0$, then such a transform is invertible,

$$f(m) = \frac{1}{C_g} \sum_{n=1}^{N} \int_0^\infty W_f(s,n)\psi_{s,n}(m)\frac{ds}{s} \tag{2.18}$$

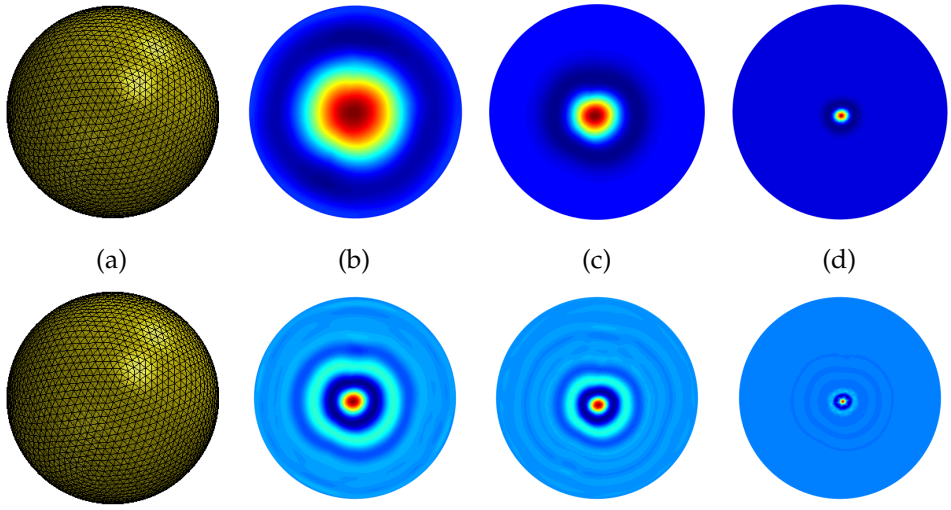


Figure 2.7: Examples of mother wavelets, i.e., Mexican hat wavelet (top) and Meyer wavelet (bottom) on a sphere surfaces. Observe that as the scale varies, the dilation of the wavelet changes. First column: A 3-D triangular mesh domain (sphere) with 10242 vertices and 20480 faces, Second column: $\psi_{1,1}$, Third column: $\psi_{2,1}$, Fourth column $\psi_{3,1}$.

which represents the original signal by superposition of wavelet coefficients and wavelet bases over the full set of scales. (2.18) is equivalent to the following expression (where we use χ_l),

$$\frac{1}{C_g}\sum_{l=0}^{N-1}\left(\int_0^\infty \frac{g^2(s\lambda_l)}{s}ds\right)\hat{f}(l)\chi_l(m). \tag{2.19}$$

which corresponds to the inverse wavelet transform in the continuous setting given in (2.6). This completes our discussion of the connection between wavelets in continuous space and a graph space.

An example of the wavelet transform on graphs discussed above is shown in Fig. 2.8 demonstrating the transformation of a random signal defined on a brain mesh. Given a brain mesh which consists of a set of vertices sampled from a brain surface and patched by triangles yielding the edges connecting the vertices, a signal (which may correspond to cortical thickness) is defined on its vertices (on the left of Fig. 2.8 left). The middle of Fig. 2.8 demonstrates

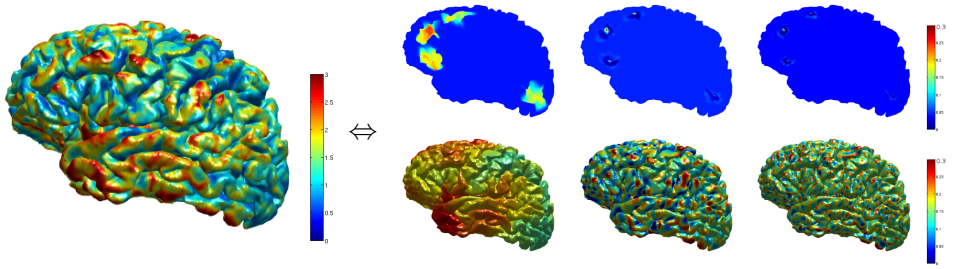


Figure 2.8: Forward and inverse graph wavelet transform on a brain mesh. Left: a random signal defined on a brain mesh, Right: examples of mother wavelets in multiple scales (top) and wavelet coefficients derived using the mother wavelets (bottom). The original signal on the left can be reconstructed using the inverse transform.

localized wavelet bases in multiple scales (top) and the resultant wavelet coefficients from (2.16) using the mother wavelets (bottom). The inverse wavelet transform using (2.18) reconstructs the original random signal on the brain mesh.

3 ANALYSIS OF CORTICAL THICKNESS USING MULTI-RESOLUTION SHAPE DESCRIPTOR

Hypothesis testing on signals defined on surfaces (such as the cortical surface) is a fundamental component of a variety of studies in Neuroscience. The goal in such an analysis is to identify regions that exhibit changes as a function of the clinical condition under study. As briefly described in Chapter 1, the clinical questions of interest move towards identifying very early signs of diseases, the corresponding statistical differences at the group level invariably become weaker and increasingly hard to identify. Indeed, after a multiple comparisons correction is adopted (to account for correlated statistical tests over all surface points), very few regions may survive. In contrast to hypothesis tests on point-wise measurements, in this chapter, we make the case for performing statistical analysis on multi-scale shape descriptors that characterize the local topological context of the signal around each surface vertex. The descriptors are based on recent results from harmonic analysis (Hammond et al., 2011), that show how traditional wavelet transform extends to non-Euclidean settings (i.e., irregular weighted graphs). In the later section, promising evidence that these descriptors successfully pick up group-wise differences is provided, where traditional methods either fail or yield unsatisfactory results.

3.1 Overview

The cerebral cortex is a layer of highly convoluted surface of gray matter with spatially varying thickness, and the distance between inner and outer cortical surface is known as the *cortical thickness*. Within the last decade, numerous studies have shown how cortical thickness is an important biomarker for brain development and disorders, and changes in the cortical thickness (Newman et al., 1998; Prevrhal et al., 1999) are particularly important in the context of Alzheimer's Disease (AD) (Erkinjuntti et al., 1987; Thompson et al., 2004; de Leon et al., 1989; Pachauri et al., 2011), which will be the primary focus of analysis in this chapter. In this context, studies have observed

significant cortical thinning in temporal, orbitofrontal and parietal regions (Lerch et al., 2005; Thompson et al., 2004) in patients with AD. Lehmann and colleagues (Lehmann et al., 2011) used both voxel-based morphometry (VBM) and cortical thickness (CT) measures extracted by Freesurfer to find significant patterns of variation between clinical populations including AD and the related posterior cortex atrophy (PCA) group. They found cortical thinning in the occipital and posterior parietal lobe in the PCA population, and in medial temporal regions in the AD population. Similar results were found in (Thompson et al., 2011; Wirth et al., 2013) which relate this measure to other biomarkers also. In many other AD studies, researchers have used cortical thickness as a biomarker to detect and classify AD cohorts from control subjects (Lerch et al., 2008; Wolz et al., 2011; Cho et al., 2012; Querbes et al., 2009; Dickerson and Wolk, 2012).

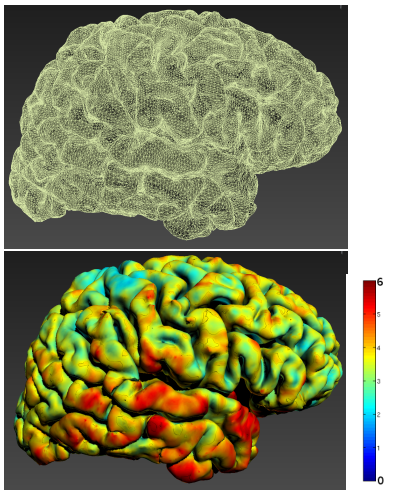


Figure 3.1: Cortical thickness measures on a brain mesh. Top: a brain surface mesh, Bottom: cortical thickness measures denoted in color.

The body of work above relating cortical thickness to cognitive decline is vast and tackles various neuroscientific questions; but these studies share a commonality in that once the thickness measurement on the cortical mesh has been calculated via a pre-processing method, the main interest is to employ statistical hypothesis testing to find regions that exhibit statistically significant differences between the two groups — typically a clinical/diseased group and a healthy control group — while accounting for various confounds.

But this workflow must take into account a few potential pitfalls. The first order requirement, clearly, is to recruit a sufficient number of subjects to ensure that the study has sufficient power. Now, if the expected variations are small, the cohort size must be large enough to ensure we can reliably identify group-wise differences. However, this may not be feasible in many cases because of several constraints due to cost, scope of clinical question and etc. Therefore, it

is important that the analysis procedure we choose is *sensitive* and maximizes the likelihood of detecting signal variations between the groups. Otherwise, in the small sample size regime, it is entirely possible that we will fail to discover a real disease-specific effect.

Notice that analysis of two very distinct groups that lie at the opposite sides of the disease spectrum will obviously yield a strong statistical signal. But recent work, with good reason, has almost entirely focused on detecting biomarkers pertaining to the *early stages* of decline (Johnson et al., 2011), or on finer gradations of the clinical spectrum from control to AD. Because of the more moderate effect size in this regime, even in larger studies, identifying group differences may be challenging. Our interest then is in deriving representation schemes for the data, which helps the downstream statistical test pick up *subtle* group differences with higher confidence than would be possible otherwise.

Multiple Comparisons. Consider the standard pipeline for analyzing cortical thickness variations in a neuroimaging study. Here, the data are defined on an intricate mesh domain (i.e., brain surface), and as a result the number of vertices needed to represent the surface (and consequently, the number of hypothesis tests) grows up to 100,000 or more. After vertex correspondences between subjects have been found, the hypothesis test is performed at each cortical surface mesh vertex. Finally, one must perform multiple comparisons correction such as Bonferroni, FDR or the method detailed in (Van De Ville et al., 2004) as the example shown in Fig. 3.2 demonstrating the effect of multiple comparisons correction (i.e., FDR) at different levels. We can then conclude that the cortical regions which correspond to the surviving vertices are indeed meaningful disease-relevant regions.

Observe that in such a vertex-wise statistical task on surfaces, improved sensitivity can be achieved by increasing the signal to noise ratio. One option may be to utilize a filtering operation (such as Gaussian smoothing). But this relies on achieving a delicate trade-off between smoothing the signal just enough to suppress noise but taking care not to blur out the signal of interest. Instead, our key idea is to derive a descriptor for each mesh vertex that characterizes its local context, at multiple scales (or resolutions) concurrently. Such

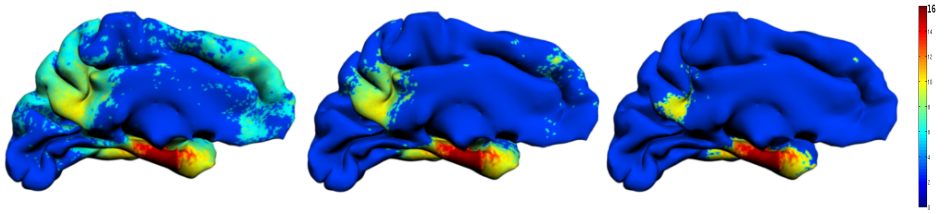


Figure 3.2: Demonstration of varying multiple comparisons correction thresholds with p-value maps on a brain surface identifying AD specific regions. FDR $\alpha = 10^{-3}$ (left column), $\alpha = 10^{-5}$ (middle column) and $\alpha = 10^{-7}$ (right column) respectively.

multi-resolution ideas, historically studied within image processing as scale space theory (Lindeberg, 1993) or via the wavelet transform (Daubechies, 1990; Mallat, 1989), have been used sparingly within the context of statistical analysis on arbitrary meshes (Chung et al., 2008; Chung, 2006). The framework presented here gives an end to end solution that makes these ideas implementable for cortical surface data, with improved sensitivity.

The wavelet transformation mentioned above and introduced in Chapter 2 is an obvious choice for multi-resolution analysis which uses a centered oscillating function as the basis instead of the sine basis. Therefore, it overcomes the key limitation of Fourier series in failing to capture sharp changes in a function (i.e., Gibbs phenomena due to infinite support) via the localization property. Unfortunately, the conventional formulation is defined only in the Euclidean space (e.g., a regular lattice). This is not suitable for convoluted and arbitrary surface models where the mesh has a highly irregular geometry. In order to still make use of the main theoretical constructs, but in the non-Euclidean setting, one must first decide a priori a "standard" coordinate system. Popular parameterization techniques use a unit sphere and utilize the spherical harmonics (SPHARM) (Chung et al., 2007). SPHARM defines Fourier bases using spherical Laplacian to parameterize a function mapped to a sphere.

The procedure above involves a module which will 'balloon' out the the cortical surface on to a sphere while preserving, to the extent possible, local distances, areas or angles. This is usually a lossy or distortion prone process (Thompson and Toga, 1996). Based on similar ideas, the spherical wavelet de-

fines the wavelet on a template sphere with discretized regular lattice (Freeden and Windheuser, 1996; Antoine et al., 2002). Studies have shown how spherical wavelets can be used to analyze complex cortical surface development (Yu et al., 2007). However spherical wavelets, like spherical harmonics, by design, cannot compensate for the metric distortion already introduced in the sphere mapping module. Of course, there are some heuristic adjustments which offer varying levels of empirical performance in practice. But theoretically, it will be satisfying to remove the restriction of a standardized coordinate system completely, and derive a multi-resolution representation in the native domain itself.

By leveraging some recent results from the harmonic analysis literature (Hammond et al., 2011) introduced in Chapter 2, we propose a framework to decompose a scalar function defined at each vertex of a surface model into multiple scales using non-Euclidean wavelets. It is easy to think of this process as viewing each mesh vertex zoomed at various levels, and characterizing the set or union of all such views within a vertex descriptor. Once such a representation is derived, we can simply analyze the multi-scale signal using multi-variate statistical tests. Later in this chapter, we show that such a framework provides substantial improvement over analyzing a uni-variate signal using cortical thickness data on brain meshes.

3.2 Multi-scale Shape Descriptor for Signals on Graphs

3.2.1 Wavelet Multiscale Descriptor

Wavelet transform is well-known for deriving multi-resolution views of signals and provides robust comparison of signals in multiple resolutions. In this sense, the fundamental idea is for using wavelet transform on graphs for cortical analysis is that we can, perhaps, achieve robust comparisons of cortical thickness measures on brain meshes which would facilitate statistical parametric mapping analyses of neuroimaging data. Using the spectral graph wavelet transform (introduced in Chapter 2) of a signal defined on a graph

yields wavelet coefficients on graphs and we can define the Wavelet Multiscale Descriptor (WMD) as a set of wavelet coefficients at each vertex n for multiple scales in $S = \{s_0, s_1, \ldots, s_{|S|-1}\}$ as

$$\mathrm{WMD}_f(n) = \{W_f(s, n) | s \in S\} \tag{3.1}$$

where s_0 denotes the scale at the scaling function. The WMD on each vertex n can be interpreted as the original univariate signal (e.g., cortical thickness,) being decomposed into various resolutions depending on the geometry of the original space. WMD is suitable for analyzing any signal defined in a non-Euclidean space (e.g., brain mesh or other 3-D shape mesh).

3.2.2 Implementation Settings

Our framework is implemented using the spectral graph wavelet transform (SGWT) toolbox from (Hammond et al., 2011) as a sub-module. First, the graph representation G of a surface mesh is derived from its Delaunay triangulation, which gives a vertex set V as well as a set of faces, each of which is comprised of a 3-tuple of vertices consisting a triangle. From these we can extract a binary edge relation E. The cortical thickness values are then computed on each vertex by Freesurfer (Fischl, 2012), which is a function $f(n)$ (or a signal) defined at each vertex $n \in V$.

In our experiments, we used the default spline wavelet design provided by SGWT toolbox as the kernel function g, which is a piece-wise function,

$$g(x; \alpha, \beta, x_1, x_2) = \begin{cases} x_1^{-\alpha} x_2^{\alpha} & \text{for } x < x_1 \\ s(x) & \text{for } x_1 \leqslant x \leqslant x_2 \\ x_2^{\beta} x_2^{-\beta} & \text{for } x > x_2 \end{cases} \tag{3.2}$$

where $s(x) = -5 + 11x - 6x^2 + x^3$, $\alpha = \beta = 1$, $x_1 = 1$ and $x_2 = 2$, which are default settings in (Hammond et al., 2011). Scales of g are defined as equally spaced bands in log scale in the spectrum of graph Laplacian. Here, the choice of the number of scales is important and must be made empirically (details below). We also note that since it is not feasible to eigendecompose a graph

Laplacian when there are more than $\sim 10^5$ vertices within a brain surface, we cannot easily access the full spectrum. What we can do instead is to find the largest eigenvalue, and then divide the spectrum into a number of bins, giving the different scales. Hence, the method has just one tunable parameter, which is the number of bins (i.e., the number of scales). This has to be a small integer, meaning that there are a very small number of values that this parameter can take, if one sets this parameter using a validation test empirically. Because noise generally lies at high end of the spectrum, we only use the scales of lower end of the spectrum to define the WMD. Those scales of interests are chosen by incrementally adding band of the scales from the coarser scale until results are satisfactory, and the remaining scales are discarded.

Since we operate on multiple brain surfaces, the range of the entire spectrum is defined by the largest eigenvalue of the graph Laplacian of all subjects. Defining wavelets in the common spectrum ensures that we define the same wavelet transform over the group of subjects. To divide up the spectrum, we ran experiments by setting the total number of scales to 5, 6 and 7. We observed empirically that 7 scales were effective for our dataset. Next, one must choose how many scales will be used to define the actual descriptor for statistical analysis. We found that using the first four scales for both datasets works reliably noting that other choices for these parameters yield comparable results.

3.2.3 Experimental Framework and Statistical Analysis

The goal of our experiments was to assess the improvement in the ability to detect group differences using WMD versus using cortical thickness on its own (i.e., baseline). The results of these experiments will be described in the following section.

We followed the general analysis pipeline (as described in Chapter 1 and shown in Fig. 1.3) by plotting the resultant p-values on the template brain surface after FDR correction. We additionally applied heat-kernel smoothing on the cortical thickness to compare the group analysis result. To compute the p-values, we used a t-test on univariate variables (i.e., raw and smoothed

cortical thickness) and Hotelling's T^2 test and Multivariate General Linear Model (MGLM) with Hotelling-Lawley trace on the multivariate variables of interest (i.e., WMD). Using MGLM, we controlled for the effects driven by factors that were not directly related to the disease (i.e., age or gender) to obtain a more accurate result.

Since our fundamental argument is that multivariate WMD is more sensitive than performing statistical tests on univariate cortical thickness, one should expect to see a stronger signal than results derived via smoothed or raw cortical thickness.

3.3 Cortical Thickness Analysis: Group Analysis for Alzheimer's disease (AD) Studies

The identification of group differences between cortical surface signals is based on comparing the distribution of the signal across the two groups at each vertex. This can be done either by using the signal (cortical thickness) obtained from the segmentation directly, or by using spherical harmonic (SPHARM) to first parameterize and then smooth the signal, followed by a vertex-wise *t*-test on the smoothed signal. In contrast to spherical approaches, our multi-scale descriptor is well defined for characterizing the shape (and the signal) on the native graph domain itself by utilizing graph structure for each mesh for signal transforms. We employ hypothesis testing using the original cortical thickness and SPHARM as the two baselines for comparison.

3.3.1 Simulation of Surface-based Group Analysis and ROC Response

We first demonstrate group analysis using WMD using synthetically generated cortical thickness (and atrophy) on a template brain surface. The template brain surface consists of 2790 vertices and 5576 faces, and 20 diseased and 20 control subjects are artificially synthesized using the template brain. First, a synthetic baseline global cortical thickness signal of mean $2mm$ and variance 0.1 is introduced. This is shown as the blue region in Fig. 3.3.

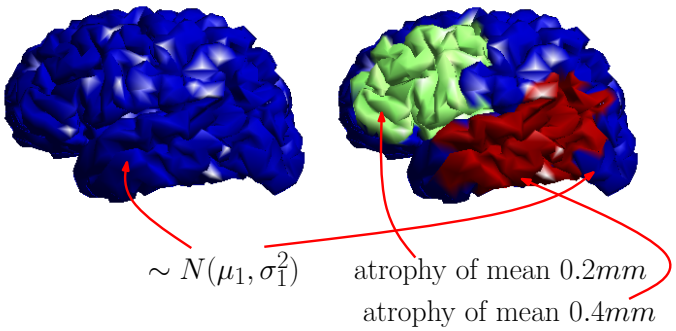


Figure 3.3: Simulation setup of synthetic cortical thickness and atrophies on brain surfaces. Blue regions correspond to the default (non-diseased) cortical thickness signal, $\mu_1 = 2\text{mm}$, $\sigma_1^2 = 0.01$. Green and red correspond to disease regions which undergo atrophy affected on the default cortical thickness signal. These atrophy levels are about ~ 25% and ~ 50% of the actual atrophies measured in AD specific regions of a real dataset.

Note that this region is viewed as not affected by disease and so no group differences should be identifiable in these regions. Next, we define two diseased regions (green and red) in Fig. 3.3. These regions undergo varying levels of atrophy (relative to the 'default' cortical thickness signal in blue). The green region corresponds to a mean atrophy of 0.2mm (variance 0.02mm) and the red region corresponds to a mean atrophy of 0.4mm (variance 0.04mm) effecting the default (blue) cortical thickness signal. The red and green regions cumulatively correspond to a total of 889 vertices (32% of the brain region). Finally, we add noise from $\mathrm{N}(0, 1)$ to the cortical thickness signal obtained from the above procedure.

From the above data, we obtain smoothed CT and WMD for comparison. Smoothing is performed via heat-kernel smoothing with bandwidth of 0.5. The spectrum of the graph Laplacian is $[0, 18.5]$, and this range is divided into 6 bins including the scaling function in order to define WMD. The spline kernel function g from SGWT toolbox is used to obtain the WMD. For statistical group analysis, a *t*-test was used for univariate raw data and smoothed data, and we used Hotelling's T^2 test for multivariate analysis for WMD. The resultant *p*-values are shown on the template surface in $-\log_{10}$ scale for comparison, see Fig. 5.3.

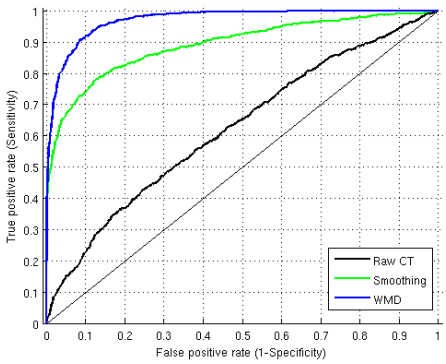
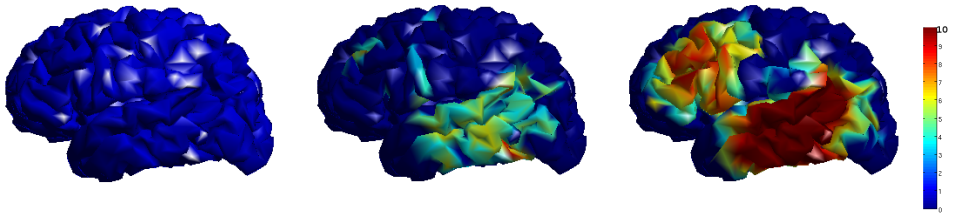


Figure 3.4: ROC curve using p-values from statistical group analysis on Raw CT, Smoothing and WMD with AUC of 0.623, 0.892 and 0.971 respectively.

It is well known that filtering raw data improves sensitivity, however, over-filtering of data may end up detecting many false positives. Multiple comparisons correction is generally applied to control the type I error in most studies. In this simulation, however, we know the ground truth from the synthetic atrophy model — a label for each vertex indicates whether it atrophies or not; so, we can conduct an ROC analysis to observe the sensitivity and specificity relationship. Here, the aim is to show that we are not only increasing the sensitivity, but we also do not make specificity worse using WMD. From this group analysis, we obtain p-values at each vertex, which tells us whether to reject the null hypothesis: the two distributions from the data at each vertex are the same. When the null hypothesis is rejected, we find those vertices with significant differences, and we can use $(1 - \mathrm{p})$ as a measure to determine the label for each vertex. The resultant ROC curve is given in Fig. 3.4, and we measure the area under the curve (AUC). We see that the raw data gives an AUC measure of 0.623, when heat-kernel smoothing is used the AUC is increased to 0.892. But using WMD yields the best AUC of 0.971 suggesting that increased sensitivity does not come at the cost of poor specificity.

Remark. Based on the simulation results, we may ask why a classical group analysis on the raw cortical thickness signal is not detecting stronger signal differences, especially since the atrophy is significant. There are two reasons for this behavior relating to the *level* of atrophy introduced in these simulations (which are not very large) and the small sample sizes. Recall that the synthetic atrophy was set to $\sim 50\%$ and $\sim 25\%$ of the mean difference in atrophy levels in disease specific regions measured in a real dataset (which was about $0.82\mathrm{mm}$). Also, the sample sizes are relatively small (20 healthy controls and 20 diseased). We will see shortly that when the atrophy differences *and* the sample sizes



(a) (b) (c)

Figure 3.5: Simulation of surface based group analysis. The resultant p-values in $-\log 10$ scale after FDR at $q = 0.01$ are mapped on the template brain surface. a) result using raw CT, b) result using heat-kernel smoothing ($t = 0.5$), c) result using WMD (4 scales out of 6).

are larger, classical analysis on cortical thickness can indeed detect regions exhibiting group-level differences.

3.3.2 Analysis on ADNI Dataset

Dataset. For the experiment, we used MR images acquired as part of the Alzheimer's Disease Neuroimaging Initiative (ADNI). Our data included brain images from 356 participants: 160 Alzheimer's disease subjects (AD) and 196 healthy controls (CN). These two groups lie at the opposite ends of the AD spectrum and the number of subjects is sufficient for group analysis. Therefore, we expect the standard method (apply hypothesis tests to the cortical thickness signal directly) to perform well and yield significant group differences. It will nonetheless provide a baseline to assess whether the multi-resolution representation yields any improvements at all. The details of the ADNI dataset are given in Table 3.1.

Experimental setup. The MRIs in ADNI dataset were pre-processed using a standard image processing pipeline, and the Freesurfer algorithm (Reuter et al., 2010) was used to segment the cortical surfaces, and calculate the cortical thickness values. In the end we obtained cortical thickness measures that were defined on cortical meshes (i.e., univariate signal defined on a graph) for each participant. We constructed WMDs for each vertex on the cortical surface at 6 different scales, and used Hotelling's T^2-test for group analysis to obtain p-values at each vertex. The resulting p-value map was corrected for multiple

Table 3.1: Demographics and baseline cognitive status measures of the ADNI dataset

Category	AD (mean)	AD (s.d.)	Ctrl (mean)	Ctrl (s.d.)
# of Subjects	160	-	196	-
Age	75.53	7.41	76.09	5.13
Gender (M/F)	86 / 74	-	101 / 95	-
MMSE at Baseline	21.83	5.98	28.87	3.09
Years of Education	13.81	4.61	15.87	3.23

MMSE: Mini Mental State Examination

comparisons over all vertices using the false discovery rate (FDR) method (Benjamini and Hochberg, 1995). The same procedure using student's t-test to acquire p-values was repeated on the original cortical thickness measurements and the smoothed cortical thickness obtained from SPHARM for baselines.

Analysis. Fig. 3.6 summarizes the results of our analysis. The first row corresponds to the result from group analysis using the original cortical thickness (CT) measures. Here, while we see some discriminative regions, group differences are weak and statistically significant in only a small portion of the brain regions. The second row shows the result pertaining to SPHARM, which indicate improvements over the baseline, partly due to the effect of noise filtering. Finally, the bottom row in Fig. 3.6 shows that performing the statistical tests using our multi-scale descriptor gives substantially larger regions with significantly lower p-values.

We evaluated the regions identified by these tests in the context of their relevance to Alzheimer's disease. We found that the identified regions are those that might be expected to be atrophic in AD. All three methods identified the anterior entorhinal cortex in the mesial temporal lobe, but at the prespecified threshold, the WMD method was more sensitive to changes in this location as well as in the posterior cingulate, precuneus, lateral parietal lobe, and dorsolateral frontal lobe. These are regions that are commonly implicated in AD, and strongly tie to known results from neuroscience (Nelson et al., 2012; Ding et al., 2014; de Leon et al., 1989).

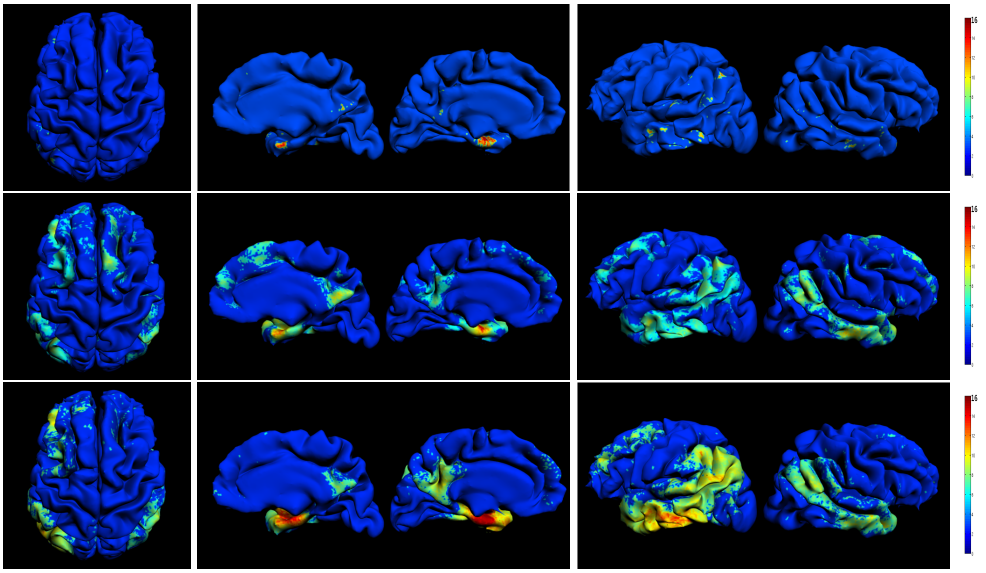


Figure 3.6: Resultant p-values in $-\log_{10}$ scale after FDR correction at $\alpha = 10^{-5}$ are shown on a template brain surface. Row 1: using original cortical thickness, Row 2: using SPHARM, Row 3: using WMD. Our framework with WMD shows improved

3.3.3 Analysis on Wisconsin ADRC Dataset

Dataset. The Wisconsin Alzheimer's Disease Research Center (W-ADRC) dataset consists of a total of 269 subjects at this point (and increasing), categorized by AD, controls, mild cognitive impairment (MCI). We used available data from 134 participants including 42 AD, 42 MCI and 50 older controls. These individuals were diagnostically characterized in the W-ADRC's multidisciplinary consensus conferences using applicable clinical criteria (McKhann et al., 1984; Petersen et al., 2001). All MCI cases were of the single or multidomain amnestic subtype whose etiology was attributed to AD. The University of Wisconsin Institutional Review Board approved all study procedures and each participant provided signed informed consent before participation.

Experimental Setup. To acquire the data, the MRI scans were obtained in the axial plane on a GE x750 3.0-T scanner with an 8-channel phased array head coil (General Electric, Waukesha, WI). 3-D T1-weighted inversion recovery-prepared spoil gradient echo scans were collected using the following parameters: inversion time (TI)/echo time (TE)/repetition time (TR)=450ms/3.2ms

Table 3.2: Demographic details and baseline cognitive status measure of the W-ADRC dataset

Category	AD (mean)	AD (s.d.)	MCI (mean)	MCI (s.d.)	Ctrl (mean)	Ctrl (s.d.)
# of Subjects	42	-	42	-	50	-
Age	76.69	10.01	75.62	7.9	76.32	5.5
Gender (M/F)	22 / 20	-	33 / 9	-	19 / 31	-
Years of Education	14.41	2.8	16.77	2.74	15.84	2.85
CDR(SB)	4.9	1.99	1.9	0.48	0.14	0.4
MMSE at Baseline	21.25	4.27	26.9	1.96	29	0.99

CDR: Clinical Dementia Rating, SB: Sum of Boxes, MMSE: Mini Mental State Examination

$/8.2\mathrm{ms}$, flip angle = 12°, slice thickness = $1\mathrm{mm}$ (no gap), field of view (FOV) = $256\mathrm{mm}$, matrix size = $256\mathrm{mm} \times 256\mathrm{mm} \times 156\mathrm{mm}$, and in-plane resolution = $1\mathrm{mm} \times 1\mathrm{mm}$. These MRIs were pre-processed using the Freesurfer algorithm as in the ADNI analysis in Section 3.3.2. Cortical thickness measures were defined on cortical meshes for each participant, WMDs were constructed for each vertex on the cortical surface at 6 different scales. Statistical group analysis was performed as in Section 3.3.2 where we compared the *p*-values from WMD and baseline (i.e., raw cortical thickness and SPHARM) analyses. On the ADRC data, we compared AD vs. controls, AD vs. MCI, and MCI vs. controls. In the AD vs. controls analysis, we expect to detect similar brain regions found in the result using the ADNI. In the AD vs. MCI and MCI vs. controls analysis, we simply show which brain regions are showing morphological changes between groups.

AD vs. Controls Analysis

We first analyze a group differences on AD and control subjects. Compared to the ADNI dataset, here we have a smaller number of subjects. Applying general hypothesis testing directly fails to detect any group differences using the raw cortical thickness due to the small sample size. Fig. 3.7 shows the resulting *p*-values in increasing order from student's *t*-test using cortical thickness and smoothed cortical thickness (heat-kernel smoothing at $t = 0.5$), and Hotelling's T^2 test and MGLM using WMD. FDR threshold at $\alpha = 0.1$ is plotted in red dotted line, and the number of vertices that are below the threshold level are considered as the signal showing significant group differences. We see that

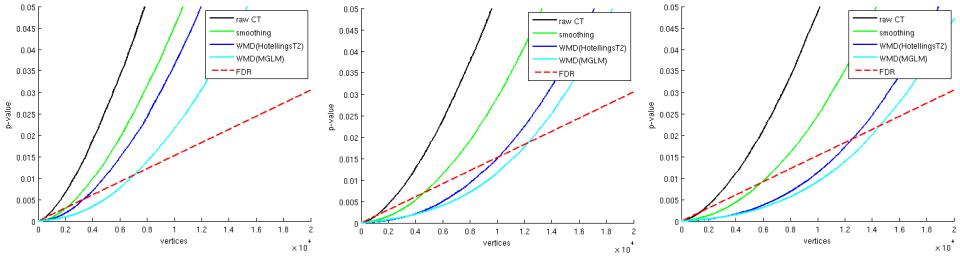


Figure 3.7: Plot of sorted p-values and FDR threshold at $\alpha = 0.1$ from AD vs. controls analysis on the right hemisphere of the brain according to different sample sizes using ADRC dataset. Left: using 60% of the total subjects, Middle: using 80% of the total subjects, Right: using all subjects. As the sample size increases, the number of surviving vertices increases. We can see that WMD increases the sensitivity.

while it is difficult to find a meaningful signal using raw cortical thickness estimates on the ADRC dataset, WMD easily detects the underlying difference. On 131076 vertices on both the right and left hemispheres of the brain surface (65538 tests on each hemisphere), we apply t-test on the cortical thickness data and heat-kernel smoothed data, Hotelling's T^2 test on WMD, and MGLM on WMD. After FDR at $\alpha = 0.05$, we detect each 622, 5913, 12455, 13769 vertices from the t-test, Hotelling's T^2 test, and MGLM respectively, which corresponds to 0.47%, 4.51%, 9.5%, and 10.5% of the total number of tests performed.

In Fig. 3.8, we compare the four different results using different features and statistical techniques on the template brain surface. In the top two rows, the result using raw cortical thickness and smoothed cortical thickness are presented. The smoothed cortical thickness helps the test detect some signal variation, but the result is weak and almost does not reveal any differential brain region. However, WMD increases sensitivity, detecting many more regions with lower p-values (using Hotelling's T^2 test); the result is shown in the third row of Fig. 3.8. Since variation of cortical thickness may be caused by age or gender, we further utilize MGLM to remove the age and gender effects. As seen in the bottom row of Fig. 3.8, the signal becomes more concentrated at specific regions.

Using cortical thickness and smoothing, we observe differences in a very small region in the right inferolateral lobe only. However, using WMD, we

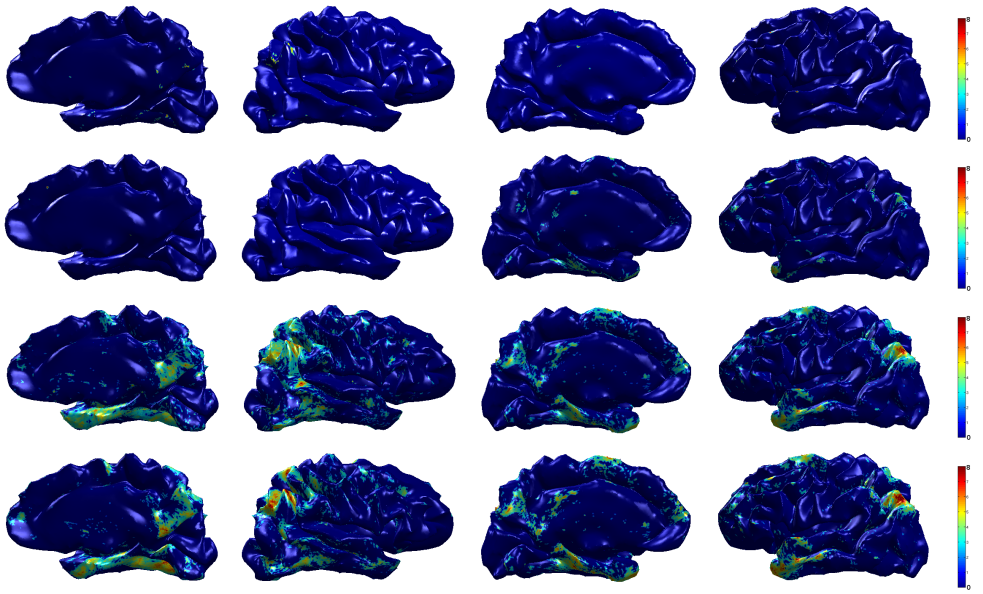


Figure 3.8: Group analysis result (AD vs. controls) on ADRC dataset. The resulting p-value(in $-\log_{10}$ scale) from hypothesis tests after FDR at $\alpha = 0.05$ is shown on a template brain surface. First row: t-test on the raw cortical thickness data, Second row: t-test on the smoothed data, Third row: Hotelling's T^2 test on WMD, Fourth row: MGLM on WMD (without age and gender effect).

find very strong group differences in the bilateral inferolateral parietal as well as temporal pole and parahippocampal cortex. Other than those regions, we also find isthmus cingulate, posterior cingulate, superior frontal, precuneus, entorhinal cortex on both right and left hemisphere as showing group differences. Since we found similar regions using ADNI dataset, it is reasonable to conclude that our results on ADNI and ADRC are in agreement.

AD vs. MCI and MCI vs. Controls

We also compare AD vs. MCI and MCI vs. controls group, and the results are shown in Fig. 3.9 and Fig. 3.10. In these analysis, we show the uncorrected p-values using Hotelling's T^2 test and MGLM (removing age and gender effect) on WMD. The first row of Fig. 3.9 and Fig. 3.10 represents the result using cortical thickness, the second row is the result using heat-kernel smoothing ($t = 0.5$), and the third and fourth row shows the result using WMD by

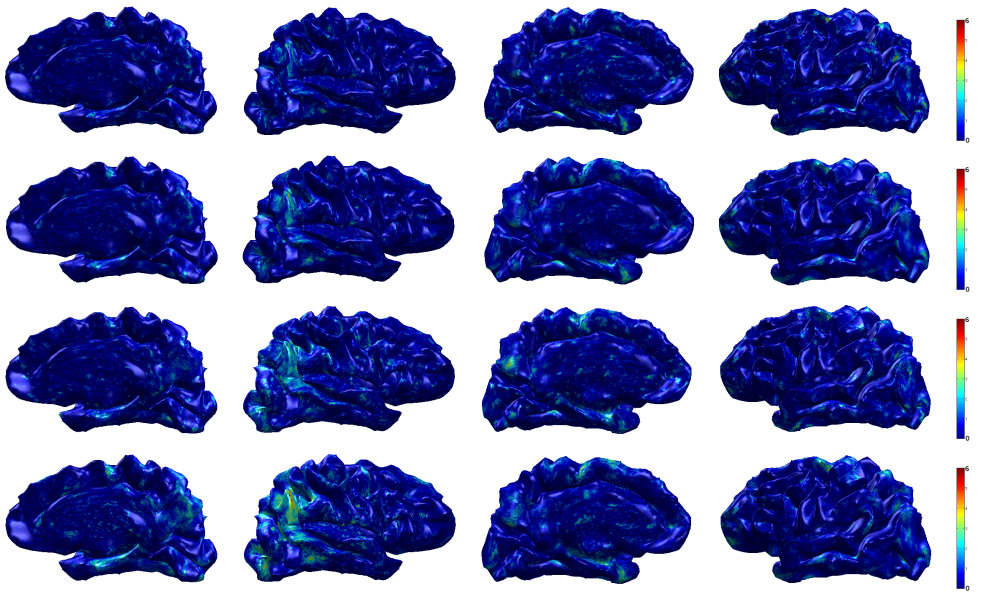


Figure 3.9: Group differences (AD vs. MCI) on ADRC dataset. First row: p-values (uncorrected) from t-test using CT, Second row: p-values (uncorrected) from t-test on smoothed data, Third row: p-values (uncorrected) from Hotelling's T^2 test on WMD, Fourth row: p-values (uncorrected) from MGLM on WMD (age and gender effect removed.)

applying Hotelling's T^2 test and MGLM respectively. These comparisons provide additional evidence that analysis with WMD is more sensitive.

In both AD vs. MCI and MCI vs. controls analysis, we expect to see similar brain regions identified by the AD vs. controls analysis, with small differences. By comparing MCI with AD and controls, we may assess the longitudinal progression of the disease in specific brain regions. On AD vs. MCI, the results showed differences in precuneus, inferior frontal and lateral occipital on both hemispheres. Relatively weaker differences in the temporal pole and parahippocampal regions are seen. As identified from the AD vs. controls analysis, the changes in cortical thickness occur in the precuneus, inferior frontal, temporal pole and parahippocampal regions as a subject enters MCI (Risacher et al., 2009; Chetelat et al., 2005). In the MCI vs. CN analysis, we observed changes in the precuneus, isthmus cingulate, inferior parietal, inferior temporal, superior temporal and temporal pole. Although not reported in

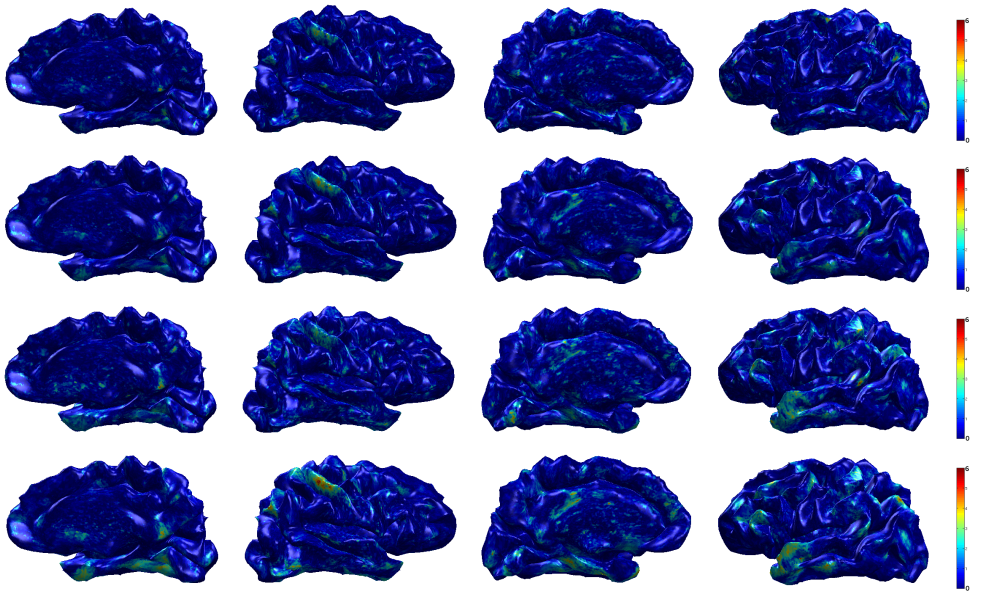


Figure 3.10: Group differences (MCI vs. controls) on ADRC dataset. First row: p-values (uncorrected) from t-test using CT, Second row: p-values (uncorrected) from t-test on smoothed data, Third row: p-values (uncorrected) from Hotelling's T^2 test on WMD, Fourth row: p-values (uncorrected) from MGLM on WMD (age and gender effect removed).

previous works, the MCI and CN comparison showed potential changes in the postcentral region as well.

3.4 Summary

Surface based mapping analysis is a widely deployed procedure in neuroimaging where we use mass univariate tests, (e.g., t-test or GLM type analysis) along with multiple comparisons correction to detect and assess statistically significant differences between clinical, genotype, or other groups of interest. The aim then is to derive maps showing the degree of significance of group level effects so as to *localize* regions of interest. This approach works very well when there are a sufficient number of subjects in the study, and when the analysis method is sensitive enough to identify such group differences. However, these assumptions may not always hold, which necessitates the design of mechanisms that are sensitive enough to identify variations even in

the smaller sample size regime. In this work, we have focused on improving the sensitivity of the extracted features, so as to mitigate the dependence on sample size.

To this end, we derived wavelet based multi-scale descriptors (WMDs) of the cortical thickness signals which are sensitive both to surface geometry and topology, as well as variations at different spatial scales. As noted above, a graph (typically) defines a non-Euclidean space, and the appropriate tools completely capture its geometry and topology. Therefore, wavelet theory lends itself nicely to the problem of deriving useful scale-dependent features. This construction is based on a set of elegant results in the harmonic analysis literature dealing with the Spectral Graph Wavelet Transform (Hammond et al., 2011) and Diffusion Wavelets (Coifman and Maggioni, 2006). This allows us to propose a multivariate approach for group analysis of surface based signals in Neuroimaging settings.

Instead of mapping the data onto a sphere, as in traditional spherical harmonic (SPHARM) based methods, our multi-scale shape descriptor is directly defined on the cortical surface graph itself, completely bypassing the ballooning process. Further, the WMD method is sensitive to signals at different scales unlike SPHARM based methods.

In our WMD construction, each scale represents a different level of support over the harmonic basis. By varying a window over the harmonic spectrum, the method of WMD efficiently characterizes both local and global context around each vertex. As the window moves toward to the lower frequency spectrum, the wavelet frame becomes more overcomplete. To avoid this issue, subsampling is utilized in grid based Euclidean spaces. However, when dealing with non-Euclidean spaces, without making any assumption on the nature of the graph, there is not necessarily a clear concept of subsampling. We note, for instance, that the method described in (Narang and Ortega, 2012) gives a method of subsampling, but only in bipartite graphs, and other methods can do this by making other assumptions. For instance, the method in (Coifman and Maggioni, 2006) assumes that the spectrum decays. Because there is no subsampling scheme on graphs, spatial correlations are induced between nearby vertices in the lower frequency range. The authors in (Van De Ville

et al., 2004) propose to deal with the spatial correlation issue by leveraging the compact, localized support of wavelets. In various situations, the above *decimated* strategy is preferable, however, the subsampling needed in such a scheme makes interpreting the specific band-pass filtering behavior difficult. Instead, the non-decimated scheme adopted here is more convenient for analysis purposes because avoiding subsampling enables deriving a descriptor precisely at the given set of vertices.

Through our experiments, we demonstrated that such a multi-resolutional shape descriptor defined in a graph space can be a powerful and flexible tool for identifying group difference signals. Indeed, a method with greater sensitivity to group differences would require recruitment of fewer subjects. We primarily evaluated the WMD framework on cortical surface signals, comparing group analysis results with WMDs against classical methods. We first compared these models using ROC curves for group analysis on synthetic brain surface data. Relative to raw uni-scale measurements, or with smoothing, we identified brain regions with much stronger group differences with global FDR correction, and in some cases these were detected when classical methods fail. In the ADNI dataset, we obtained pronounced group differences in the anterior entorhinal cortex, posterior cingulate, precuneus, lateral parietal lobe and dorsolateral frontal lobe. Similar regions were found using the distinct W-ADRC dataset as well. It is encouraging that these independent results are in agreement; we also note that these regions are consistent with the literature (Lehmann et al., 2011; Thompson et al., 2011; Wirth et al., 2013). In addition to these results, we applied MGLM analysis to control for factors such as age and gender and show how these factors change the results. The FDR-curves suggest that up to twenty-eight times more vertices using WMD than using raw cortical thickness can survive global correction. Using the W-ADRC dataset, we further analyzed the effect from MCI. We showed that WMDs obtain lower p-values than raw cortical thickness, and displayed whole-brain sensitivity map using the R^2 metric of effect size. Finally, the power analysis on AD and controls using cortical thickness and WMD indicates that WMD is more sensitive, giving smaller sample size estimates. By applying our framework on two different datasets: the ADNI dataset (a large and well characterized dataset) and the

W-ADRC dataset (central to a number of local studies,) we have demonstrated that the methodology is broadly applicable.

We believe, for several reasons, that the improved sensitivity is attributable to the filtering effect achieved by separating high-frequency information from low-frequencies. First, anatomical brain features, and neurodegenerative morbidity effects, tend to exhibit a certain degree of spatial cohesion and locality (Braak and Braak, 1995; Hinrichs et al., 2009). In addition, most noise processes, whether derived from scanner effects or post-processing, tend to be distributed across all scales. Gaussian smoothing and filtering is therefore quite common for this and other reasons. Note however, that a non-adaptive Gaussian blur kernel is oblivious to anatomical divisions such as sulci and cortical boundaries, and may inappropriately mix signals which are close spatially, but not anatomically. Heat-kernel smoothing attempts to resolve this issue by first expanding the cortices to a spherical surface (ballooning) and then smoothing, but in doing so it smooths all scales with the same fixed-bandwidth kernel. A key feature of the WMD approach is that each scale corresponds to a particular band-pass filter in the spatial domain, which can be thought of as smoothing only certain frequencies. In graph-based methods, smoothing can, and indeed must, be done separately for each scale because there is a strong dependence on the unique topology of each subject's cortical surface mesh. In the interest of space, we do not report the effect of using all seven scales, (as opposed to treating the upper three as "high-frequency" signal and discarding them as noise,) but briefly, doing so uniformly weakened results and lessened significance. Moreover, the high frequency components simply did not correspond with any identifiable brain regions, and visually resembled a random "speckle" pattern. This is an important observation because in some image processing domains, high-frequency information can give well-defined edges, but this did not appear to be the case in this application.

Although we have demonstrated that our framework is able to obtain strong and robust results in group analysis, there are nevertheless a few shortcomings. Our method leaves it to the user to define the scales, and we note that this is often the case in wavelet-based methods. Ideally, one would like to eigendecompose the entire graph Laplacian, and divide the spectrum into portions

of roughly equal mass. However, when there are $\sim 10^5$ nodes in the graph, this becomes infeasible. A more practical approach is to find the largest eigenvalue, and simply divide the spectrum into a fixed number of equal-width bins, which is the approach we followed. It still remains to choose how many such bins to use, but we found empirically that a small number, on the order of five to ten, works well. This choice is driven by several considerations. Primary among these is computational burden. Consider that WMDs contain information not only about the function defined on vertices, but also about the distinct topology of each subject's vertex mesh. Therefore, resampling to a grid must be done subsequent to any calculation of WMDs, and it must be done *independently* for each scale. For a large number of scales this cost becomes a bottleneck. Moreover, we do not wish to incur the *curse of dimensionality* any more than is necessary. That is, while multi-resolution descriptors can effectively separate out some signals that are scale-dependent, if we allow the descriptors to unduly proliferate then we may dilute the underlying signal by spreading it too thinly over a large number of scales. Thus, while some signals genuinely exist only at a particular scale, if we choose too many scales in some neighborhood of the true signal, then this signal may "leak" between them due to sampling artifacts. Taking these issues into consideration, we avoid choosing too many scales. In our evaluations, we found that seven scales gives satisfactory results, though we did not perform an exhaustive grid search because of the above mentioned constraints. In addition to the choice of number of bins is the choice of which ones to discard as high-frequency noise, and which to treat as low-frequency signal. Following a similar line of reasoning as above, we simply chose the first four bins as *signal* and the last three as *noise*. This is corroborated by the fact that the distribution of p-values of the WMDs corresponding to high frequency portions of the spectrum followed a roughly uniform distribution, and visual inspection showed no recognizable spatial cohesion. This is exactly as we expect, and is in fact the intended effect — considering that the cortical thickness measures can be recovered as a deconvolution and summation of the WMDs, and that the high frequency WMDs are designed to serve as a model of "noise", then it is not surprising that the overall cortical thickness signal is weaker.

There is an *important* increase of interest in wavelet based neuroimaging analysis methodologies. A number of works have expanded the basic framework as well as adapted it to various statistical issues. Van De Ville and colleagues (Van De Ville et al., 2004) approach the problem of selecting thresholds for both wavelet and spatial domains, which is important because without addressing this issue, spatial statistical maps are uninterpretable. They tackle this issue by balancing the two thresholds in the wavelet and the spatial domain, and apply statistical testing in the spatial domain instead of the wavelet domain. The authors of (Leonardi and Van De Ville, 2013) explored the ramifications of using tight (or Parseval) wavelet frame which more closely resembles an orthonormal basis while retaining basic wavelet properties. Among the advantages of using a tight frame is that the inverse transform is easy to compute; this plays an important role in pre-processing the raw data. In addition, it is efficient because it preserves the energy in the transformed domain. Although the tight frame formulation is preventing spectral leakage between scales and proposes a much cleaner strategy than re-normalizing the coefficients, the t and T^2 statistics we used in the analysis automatically re-scale the coefficients regarding the difference in ranges between scales. The authors above also noted that the construction in (Hammond et al., 2011) does not give a tight frame, however it is nonetheless feasible for our particular application (as shown in our experiments) because we are primarily interested in the forward transform which is essential to obtaining the descriptors.

Lastly, we observe that there is an issue of whether to account for subject specific variations in global cortical thickness. Adjustments for global effects in volumetric analyses are premised on the finding that individuals who overall have bigger heads also tend to have larger regional brain structures (e.g., hippocampus) than persons with smaller heads. Normalization of regional volumes by means of whole brain volume (whether via regression approaches or proportional scaling) is therefore necessary to control for this potential confound. In contrast, available evidence convergently indicates that cortical thickness is only minimally or not at all related to sex, height, or overall brain size. Therefore, adjusting for brain size/global thickness while performing vertex-wise cortical thickness analyses risks introducing error variance into the

model. (Salat et al., 2004; Dickerson et al., 2009; Palaniyappan, 2010; Whitwell et al., 2013)

Despite these unresolved issues, our results shown in this chapter suggest that the method may be highly suitable to traditional group analysis in most cases. The procedure can be easily adapted to analyze data with arbitrary topologies (Kim et al., 2012; Chung et al., 2005) and for studies dealing with other neurodegenerative disorders involving morphological measurement on the brain surface or on brain networks as we will describe in later chapters.

4 MULTI-RESOLUTION DESCRIPTOR ON BRAIN CONNECTIVITY

There is significant interest, both from basic and applied research perspectives, in understanding how structural/functional connectivity changes can explain behavioral symptoms and predict decline in neurodegenerative diseases such as AD. The first step in most such analyses is to encode the connectivity information as a graph; then, one may perform statistical inference on various 'global' graph theoretic summary measures (e.g., modularity, graph diameter) and/or at the level of individual edges (or connections). For AD in particular, clear differences in connectivity at the dementia stage of the disease (relative to healthy controls) have been identified (Sheline et al., 2010; Shao et al., 2012).

Despite such findings, AD-related connectivity changes in preclinical disease remain poorly characterized. Such preclinical datasets are typically smaller and group differences are weaker. In this chapter, we propose a new multi-resolution method for performing statistical analysis of connectivity networks/graphs derived from neuroimaging data. At the high level, the method occupies the middle ground between the two contrasts — that is, to analyze global graph summary measures (global) or connectivity strengths or correlations for individual edges similar to voxel based analysis (local) (Bullmore and Sporns, 2009). Instead, our strategy derives a wavelet representation at each primitive (connection edge) using the wavelet transform on graphs introduced in Chapter 2 which captures the graph context at multiple resolutions. We provide extensive empirical evidence on how this framework offers improved statistical power by analyzing two distinct AD datasets. Here, connectivities are derived from diffusion tensor magnetic resonance images (DT-MRI) by running a tractography routine. We first present results showing significant connectivity differences between AD patients and controls that were not evident using standard approaches. Later, we show results on populations that are not diagnosed with AD but have a positive family history risk of AD where our algorithm helps in identifying potentially subtle differences between patient groups.

4.1 Overview

Alzheimer's Disease (AD) is a progressive neurodegenerative condition characterized by severe loss of cognitive function and ability to carry out activities of daily living (McKhann et al., 2011). It has a long course and significant pathology can accumulate prior to development of clinically relevant cognitive impairment (Perez-Nievas et al., 2013; Chételat et al., 2013). While analyses of regional brain changes have come a long way toward informing upon the preclinical stages of AD, it has become increasingly clear that a better understanding of AD may not be possible based on characterizing regional pathology alone. AD poses a challenge given that the cognitive changes that define the disease do not manifest until significant brain pathology has accumulated, and because cognitive changes are poorly correlated with certain pathological features of the disease such as amyloid accumulation (Jack et al., 2013). A promising initiative then, is to derive a better understanding of the disease by characterizing changes in connectivity, taking into account the neural networks that comprise several affected regions. Functional connectivity changes are well-documented in the disease (Wang et al., 2007; Damoiseaux et al., 2012; Supekar et al., 2008), and promising studies suggest that connectivity changes can explain behavioral symptoms in AD and may predict conversion to AD (Filippi and Agosta, 2011; Li et al., 2002; Shao et al., 2012). Despite clear differences in connectivity at the dementia stage of the disease, AD-related connectivity changes in preclinical disease are not well-characterized. In this chapter, we use a novel approach to probe connectivity differences in a cohort of late-middle-aged adults enriched for risk factors for AD, including parental family history and APOE $\epsilon 4$ genotype.

Our approach builds upon prior studies using graph-based network analysis to assess connectivity. Assume we are given a population of connectivity graphs, $G = \{G_i\}$ where $i \in \{1, \cdots, n\}$ indexes the study participants. Each graph $G_i = \{V_i, E_i\}$ corresponds to a single subject, the vertex set V_i is an anatomical starting point, and each edge in the set E_i provides information about the relationship between these vertices. In this study, connectivity was based on the strength of white matter tract connections between template-

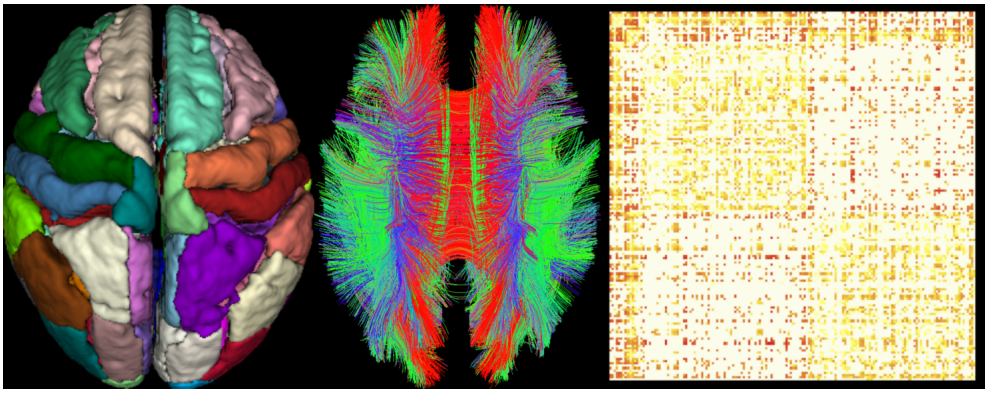


Figure 4.1: Deriving connectivity matrix using tractography. Left: Brain atlas, Middle: fiber tracts from DTI, Right: connectivity matrix derived from the atlas and fiber tracts. Each element in the connectivity matrix represents strength measures of tract bundles between ROIs defined by the atlas.

defined anatomical locations, derived from diffusion tensor imaging (DTI). Each voxel in DTI gives the orientation of water molecules in the brain which gives us directional information given as a tensor at each voxel location. By following these tensors, we obtain individual neural tracts in the brain (middle in Fig. 4.1), and by selecting those tracts connecting different brain atlas (left in Fig. 4.1), we identify tract bundles connecting different regions in the brain. In the end, we get a connectivity matrix (right in Fig. 4.1) by defining a strength measure for each track bundle such as FA, MD or the number of fibers as we introduced in Chapter 1.

In order to improve sensitivity, we employed multi-resolution analysis using wavelets described in Chapter 2. Multi-resolution analysis exploits the "local context" of information to identify significant effects. That is, multi-resolution analysis takes into account the fact that information can be viewed at different resolutions (similar to zooming in and out of an image), which provides a mechanism to capture the 'context' of information when performing the downstream statistical analysis. To see why such an approach may be useful, observe that roughly speaking, the connectivity analysis literature may be broadly clustered in two categories. The first set of approaches take into account graph theoretic summaries of the entire graph (i.e., girth, diameter,

modularity, small-worldness, degree distribution, etc.). Alternatively, we may go with more local approaches, i.e., a Voxel Based Analysis (VBA) type analysis applied to graph edges (Fornito et al., 2013; Bullmore and Sporns, 2009). To obtain disease specific behavior of specific edge connections, we see a need for a framework that lies in the continuum between the *global* and *local* methods. These requirements make wavelets a promising way to approach this problem. This approach is based on the concept that a signal can be represented by a set of localized wavelet coefficients, while noise in the signal is uniformly spread throughout the wavelet space (Ruttimann et al., 1998). Recent advances in multi-resolution wavelet analysis have facilitated the development of sensitive methods for image analysis, including the recent work on wavelet based morphometry (Canales-Rodríguez et al., 2013) in neuroimaging. The technique offers certain advantages in that it improves sensitivity and specificity relative to voxel based analysis, arguably due to the multi-resolution perspective of the data (Canales-Rodríguez et al., 2013). Tract based spatial statistics (TBSS) may also be considered a hybrid global/local approach (Smith et al., 2006). We provide a more detailed discussion of this issue later in Section 4.5 and relate it to the ideas developed here.

While multi-resolution wavelet analysis has been applied more extensively in classical image processing where images are signals sampled on a uniform lattice, connectivity graphs are not. The standard constructions are no longer applicable for non-Euclidean spaces. Based on recent work in harmonic analysis on spectral graph wavelets (Hammond et al., 2011) and using methods described in Chapter 3 to apply non-Euclidean wavelet based transformations to conduct shape analysis (Kim et al., 2012, 2014), we show how to perform multi-resolution wavelet analysis to connectivity graphs derived from DTI data. Multi-resolution wavelet analysis is ideal for improving sensitivity in both cases where sample sizes are low (often the case in patient-based studies), and where differences (effect sizes) may be small, often the case in studies of preclinical participants.

We evaluated preclinical alterations to structural connectivity in a late-middle-aged group of adults from the Wisconsin Registry for Alzheimer's Prevention (WRAP) study. In order to limit our analysis to white matter

tracts that change in known disease, we first applied our approach to AD patients compared to age-matched controls from the Wisconsin Alzheimer's Disease Research Center (ADRC). The starting point for tract connectivity was information derived from fractional anisotropy (FA), a summary measure of directional water diffusion that is highly sensitive to microstructural features including axonal density, diameter, myelination, and cytoskeletal features. We hypothesized that non-demented adults with increased risk for AD due to parental family history, would show differences in connectivity compared to controls. Further, given that our proposed approach is new, we compared it against standard methods of analysis (i.e., applying statistical test at all edges and correcting for multiple comparison problem), and hypothesized that we would find improved sensitivity using multi-resolution analysis.

4.2 Line Graphs of Connectivity Networks

As mentioned in Chapter 2, a graph G consists of a vertex set V, an edge set E and corresponding edge weights ω. In order to utilize a graph wavelet transform to derive a multi-resolution view of the signals defined on the graph edges (i.e., edge weights ω), one must first define the edge weights as signals defined on a new graph. For this, a dual representation of a graph G, the line graph $L(G)$, can be defined by simply exchanging the notion of nodes and edges; a node/vertex v_i with two incident edges e_{ij} and e_{ik} is now thought as if the node were connecting the two edges. After this line graph transformation, these two edges become two distinct vertices in $L(G)$, and they are connected by a binary edge (corresponding to the common vertex v_i in G).

Formally, a line graph $L(G) = \{V_L, E_L, \omega_L\}$ is defined by a vertex set V_L, an edge set E_L, and a corresponding edge weights ω_L, which is derived from the initial graph G. The vertex set V_L and the edge set E_L come from the edge set E and the vertex set V in G respectively, where as the edge weights ω_L for the edges in E_L are binary — the weight is 1 if there exists a common vertex between two edges, and 0 otherwise (Harary, 1969). It is easy to see that $L(G)$ has N_L vertices, same as the number of edges in G. One benefit of this transformation is that the original edge weights ω from G can now be treated

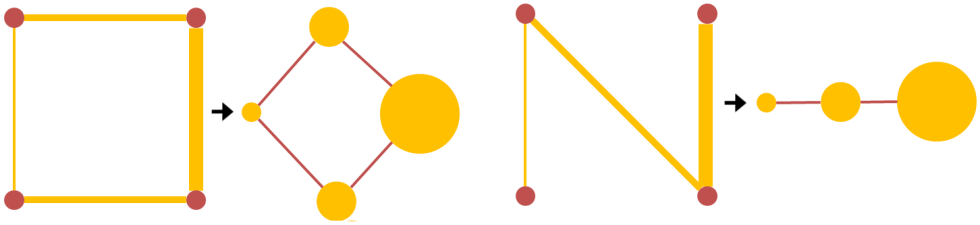


Figure 4.2: Examples of simple graphs and the corresponding line graphs. The original graphs are shown with the vertices in red and the edges in yellow, and the thickness of edges represents the edge weights, and corresponding line graphs are shown with yellow vertices and red edges. The vertex size is proportional to the signal defined on each vertex (i.e. the edge weight of the original graph).

as a function defined over the vertices V_L in $L(G)$. Consider the adjacency matrix A_L of $L(G)$, of size $N_L \times N_L$, comprised of entries g_{ij} defined as,

$$g_{ij} = \begin{cases} 1 & \text{if } v \in V,\ v \backsim e_i, e_j \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

where v is a vertex from V and e_i and e_j are two different edges in E. This means that when the ith edge e_i and the jth edge e_j are connected by a common vertex v, we assign 1 to g_{ij}, otherwise zero. Since A_L gives the connectivity between edges through vertices, it is also known as the edge-adjacency matrix. Simple examples of the line graph transformation are shown in Fig. 4.2. Here, the original graphs are represented by red vertices and yellow edges where the edge thickness denotes the edge weights. On the other hand, the line graphs consist of yellow vertices and red edges, where the vertex size represents the edge weight.

The advantage of using line graph transform is that we can now adopt edge weights in the original graph as a function defined on vertices in the line graph. Once the line graph is obtained from an original graph, the edge weights in the original graph become a function that is defined on the vertices of the transformed line graph. We can then apply multi-resolution analysis using wavelets on the function defined on this line graph.

4.3 Multi-resolution Descriptor on Graph Edges

Recall from Chapter 2 that the wavelet transform starts by defining a mother wavelet $\psi_{s,a}$, which is a basis function with scale (s) and translation (a) properties. For a graph where the vertices are indexed by n, a wavelet transform of a signal $f(n)$ using $\psi_{s,n}$ as basis is defined as

$$W_f(s,n) = \langle f, \psi_{s,n} \rangle \tag{4.2}$$

to obtain the wavelet coefficients $W_f(s,n)$. Now that we can define the graph edge weights ω as a function defined on the nodes of $L(G)$, we can obtain wavelet coefficients from the ω on $L(G)$ using spectral graph wavelet transform (Hammond et al., 2011) in the same manner as in Chapter 3. Here, the $W_f(s,n)$ in scales of $S = \{s_0, s_1, \ldots, s_{|S|-1}\}$ are defined at each node of a line graph. When they are transformed to the original graph domain, they form the Wavelet Connectivity Signature (WaCS), a multi-resolution descriptor on each edge of the original graph as

$$WaCS_f(e) = \{W_f(s,e) | s \in S\} \tag{4.3}$$

where e is a edge index from the original graph which corresponds to the node index of the line graph. Notice that e and n are equivalent here and therefore the WaCS is an analog of WMD (in Chapter 3), but defined on graph edges. Since the line graph transformation maps the original connectivity graph to the dual representation where each vertex has an associated signal measurement, the wavelet coefficients will provide a multi-resolution view of the connectivity signal.

4.4 Brain Connectivity Discrimiation: Group Analysis for AD and Preclinical AD

In this section, we provide experimental results from the analysis of AD and a AD-related risk factor such as family history (FH). Here, family history positive (FH+) is defined where any of the participant's biological parents have been

diagnosed as AD and family history negative (FH-) otherwise. The participants went through diffusion tensor imaging (DTI) scans and tractography process was performed on the DTI to obtain structural brain connectivity for each participant. We use two independent datasets to setup two-stage experiments, 1) Alzheimer's Disease Research Center (ADRC) dataset which consists of participants in either AD or controls group and 2) Wisconsin Registry for Alzheimer's Prevention (WRAP) where the participants are healthy controls but at risk of AD. The analysis consists of two stages, 1) AD versus controls and 2) FH positive versus FH negative analyses, where we perform statistical group analyis at each brain connectivity. We first analyzed AD symptoms using ADRC dataset and derive connections of interest (COI). In the second stage, we used WRAP datset to perform group analysis at each COI to localize the affect of FH conditioned on the AD related variations. The details of the experiments and results are described in detail in the following sections.

4.4.1 Deriving Structural Brain Connectivity from Diffusion Tensor Imaging

DTI acquisition: Participants were imaged on two identical General Electric 3.0 Tesla Discovery MR750 (Waukesha, WI) MRI systems fitted with an 8-channel head coil and using parallel imaging (ASSET). All participants in the W-ADRC dataset were imaged on one scanner, while all WRAP participants were imaged on a second, identical scanner. For both cohorts, DTI was acquired using a diffusion-weighted, spin-echo, single-shot, echo planar imaging (EPI) pulse sequence in 40 encoding directions at $b = 1300 s/mm^2$, with eight non-diffusion weighted ($b = 0$) reference images. The cerebrum was covered using contiguous 2.5 mm thick axial slices, FOV = 24 cm, TR = 8000 ms, TE = 67.8 ms, matrix = 96×96, resulting in isotropic 2.5 mm^3 voxels. High order shimming was performed prior to the DTI acquisition to optimize the homogeneity of the magnetic field across the brain and to minimize EPI distortions.

Image analysis: We employed a robust processing pipeline, based on methods in (Zhang et al., 2007a) and reported in (Adluru et al., 2014). The processing stream is depicted in Fig. 4.3. For both datasets, head motion

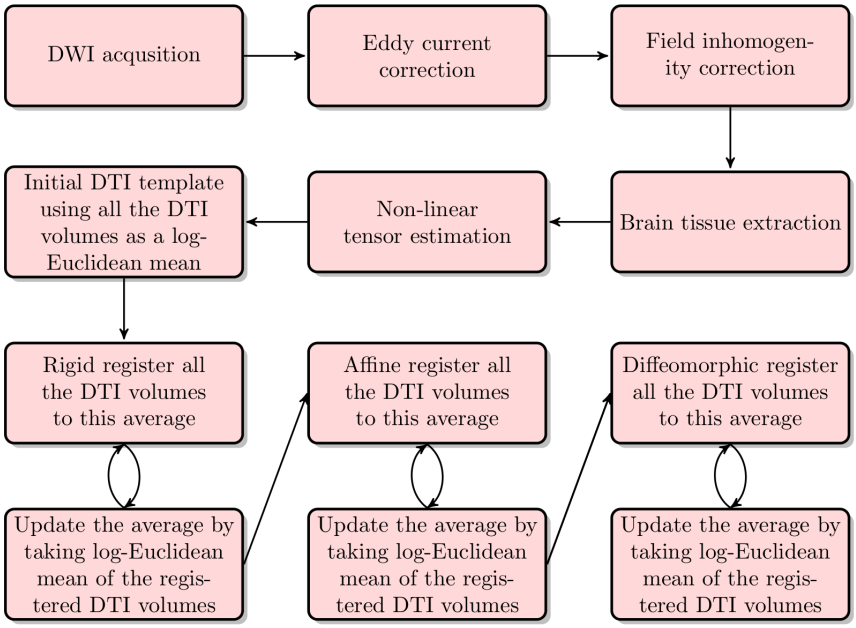


Figure 4.3: The pipeline for generating the template: After the data are acquired, the DWI images are corrected for eddy current distortions and field inhomogeneity. Then, brain tissue is extracted from the images so further processing is done only on the relevant regions of the images. Tensors are estimated by non-linear optimization. An initial bootstrap template is then computed using the Log-Euclidean mean approach. Finally, the bootstrap template is iteratively improved using three layers: rigid registration, then affine registration and lastly by diffeomorphic registration.

and image distortions (stretches and shears) due to eddy currents were corrected with affine transformation in the FSL (FMRIB Software Library) package (`http://www.fmrib.ox.ac.uk/fsl`). Geometric distortion from the inhomogeneous magnetic field applied was corrected with the $b = 0$ field map, PRELUDE (phase region expanding labeler for unwrapping discrete estimates) and FUGUE (FMRIB's utility for geometrically unwrapping EPIs) from FSL (Smith et al., 2004), where field maps were available. All images were visually inspected at this stage to ensure that data with substantial artifact (loss of frontal or temporal lobe signal) or geometric distortions are not included in the final analysis. Brain tissue was extracted using FSL's BET (Brain Extraction Tool). Tensor fitting was performed using a nonlinear least squares method in

Camino (Alexander and Barker, 2005).

Spatial normalization: For each of the dataset, we first created population specific templates using Diffusion Tensor Imaging ToolKit (DTI-TK, `http://www.nitrc.org/projects/dtitk/`) which is an optimized DTI spatial normalization and atlas construction tool (Zhang et al., 2006a, 2007b) that has been shown to perform superior registration compared to scalar based registration methods (Adluru et al., 2012). The template is constructed in an unbiased way that captures both the average diffusion features (e.g., diffusivities and anisotropy) and the anatomical shape features (tract size) in the population (Zhang et al., 2007b). Individual maps were then registered to the study-specific templates using rigid, affine, and diffeomorphic alignments and finally interpolated to $2 \times 2 \times 2$ mm^2, voxel resolution before generating the network data. Note that we do not use an overall template representing both datasets since the node regions of interest provide us the correspondence across subjects and datasets.

Tractography based network estimation: Tractography was performed on the template using the following parameters for the `track` command implemented in Camino (Cook et al., 2006): curvature threshold of 45°, curve interval (i.e., the length of tract over which the curvature threshold is tested) of 10 mm. An Euler tracking algorithm with nearest neighbor interpolation based on probabilistic tensor deflection (TEND (Lazar et al., 2003)) was used to generate the tracks, and the step size for the Euler tracking was set to 0.1 mm. Bootstrapping (20 times) was performed (i.e., 20 tracts were generated per seed voxel where the seeds were defined as the entire brain mask.) The seed mask was obtained by binarizing the trace map of the population template where the lower and upper thresholds were 0.01 and 100 respectively. These thresholds are commonly used in DTI-TK (Zhang et al., 2006a). The stopping criteria did not involve any thresholding for FA since the tracts were filtered to pass through the ROIs from the IIT atlas. Note that the bootstrap tractography procedure was used *only* to define the white matter region of interest which connects two separate nodes. Once the regions were defined from the tractography procedure, we took the mean FA values along the tracts, which is stable and is not affected by the number of bootstrap replications.

The number of replications (20) was chosen based on our initial experiments to make sure we obtained reasonable white matter pathways between pairs of nodes and verified that changing the number of replications has no effect on the white matter region extracted.

Then, a DTI white matter atlas (Varentsova et al., 2014) was registered to the population average FA map using Advanced Normalization Tools (ANTS) (Klein et al., 2009). In total, 164 regions were initially defined on a diffusion tensor template which has been made publicly available (Varentsova et al., 2014). Once we warped the FA of that template to the FA of the population template, we applied the warp to the 164 regions to align them onto the population template using nearest neighbor interpolation. In the end, two regions labeled `unknown` were excluded. Full brain tractography was then used with the `conmat` tool in Camino to obtain edge weights using the spatially normalized FA and MD measures giving us two 162×162 symmetric matrices per subject. A full index of the ROIs can be found in IIT3 atlas documentation at `http://www.nitrc.org/projects/iit2`.

4.4.2 Brain Connectivity Differences between AD and Controls

Dataset. For our AD versus control analysis, we used the Wisconsin Alzheimer's Disease Research Center (W-ADRC) dataset, which we used in our experiment earlier in Chapter 3. In this experiment, we use the data from $N = 102$ participants who went through DTI scan. The demographic of the W-ADRC dataset used in this analysis is given in Table 4.1. Each brain network in the dataset is given as a 162×162 matrix where each element represents fractional anistropy (FA) values of neuron fibers connecting different brain regions.

Analysis. The baseline analysis using standard procedure (t-test and multiple comparisons correction) on the FA weighted edges initially yielded 6 connections. On the other hand when we performed the test (using multivariate approach) on WaCS, we detected 81 out of 13041 brain connections spanning over 67 brain regions. For the baseline approach, the p-values were computed using a general linear model and a multivariate general linear model was used to obtain the p-values from WaCS. In both cases, age and gender were

Table 4.1: Demographics of W-ADRC dataset

Category	AD(mean)	AD(s.d.)	CN(mean)	CN(s.d.)
Number of subjects	44	-	58	-
Age	77.05	9.35	74.05	6.82
Sex (M/F)	31 / 13	-	33 / 25	-
Global CDR	0.74	0.37	0.06	0.16
MMSE	22.72	4.44	29.15	0.8543
RAVLT total raw score	20.52	8.07	43.76	8.49
RAVLT long delay raw score	0.66	1.46	8.41	3.12

CDR: Clinical Dementia Rating, MMSE: Mini Mental State Examination,
RAVLT: Rey Auditory Verbal Learning Test

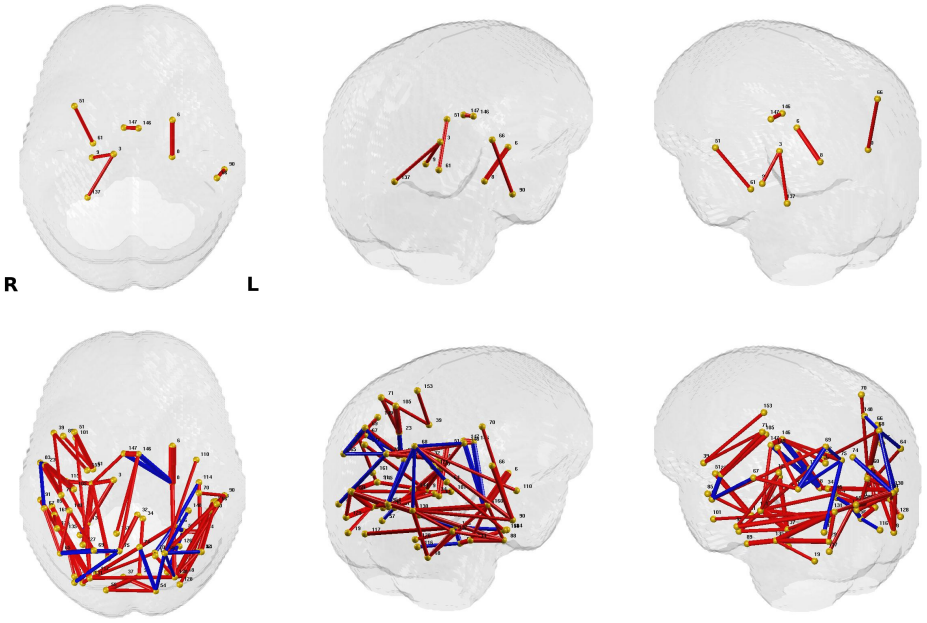


Figure 4.4: Significant group differences (controlled for age and sex) after Bonferroni correction at $\alpha = 0.01$ from AD vs. control connection analysis using FA. The thickness of each connection represents the p-values in $-\log 10$ scale (thicker connection corresponds to lower p-value), and the color of each connection represents the direction of the difference (red: stronger in controls group, blue: stronger in AD group). Top row: result from GLM on raw FA values, Bottom row: result from MGLM on WaCS and Bonferroni at $\alpha = 0.01$.

used as nuisance covariates. To keep comparisons fair, both these analyses were corrected for multiple comparisons using the Bonferroni correction at

$\alpha = 0.01$. Among the 81 connections identified with our framework, 19 of them showed higher FA in the AD group compared to control, while 62 connections showed a reverse phenomenon, i.e., higher FA values in the control group. These connections surviving the Bonferroni threshold are visualized in Fig. 4.4, the connections from baseline approach are shown in the top row and those from our framework are presented in the bottom row. Here, the thickness of each connection represents the p-values in $-\log 10$ scale (thicker connection corresponds to lower p-value), and the color of each connection represents the direction of the difference (red: stronger in controls group, blue: stronger in AD group).

4.4.3 Effects of Family History Risk on Brain Connectivity

Dataset. For our second stage analysis, we used dataset from the Wisconsin Registry for Alzheimer's Prevention (WRAP) study. The WRAP dataset consists of healthy individuals only, but categorized by the presence or absence of certain AD risk factors. The dataset contains various types of brain scans such as MRI, DTI, FDG-PET and PiB-PET on the participants under study, and will be used not only in this chapter but also in Chapter 5 and 6. The demographics of the WRAP dataset used in this experiment are given in Table 4.2.

Table 4.2: Demographics of WRAP dataset

Category	FH-(mean)	FH-(s.d.)	FH+(mean)	FH+(s.d.)
Number of subjects	93	-	250	-
Age	62.96	5.84	60.29	6.89
Sex (M/F)	64 / 29	-	166 / 84	-
MMSE	29	3.26	29.25	2.16
RAVLT total raw score	50.58	10.20	51.08	8.70
RAVLT long delay raw score	10.63	2.96	10.77	2.69

CDR: Clinical Dementia Rating, MMSE: Mini Mental State Examination,
RAVLT: Rey Auditory Verbal Learning Test

It is obvious that the differences in healthy controls caused by a disease risk factor are expected to be subtle, making the task of finding group differences a more challenging task. We therefore selected connections of interest (COIs) based on the AD versus control analysis described above (section 4.4.2), and

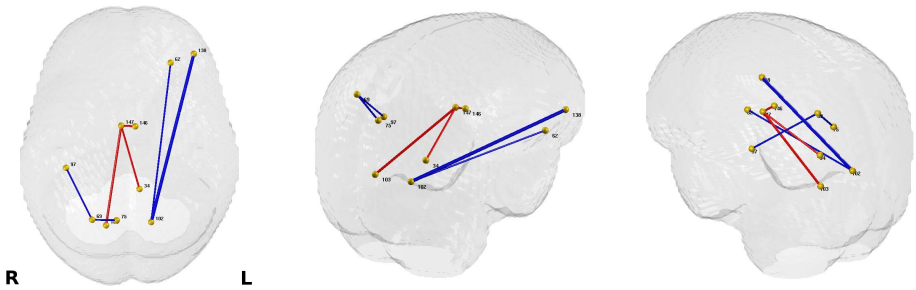


Figure 4.5: Significant group differences (controlled for age and sex) from FH+ vs. FH- connection analysis using WaCS derived from FA. Those connections with p-values that survive FDR threshold at 0.05 are exhibited. The thickness of each connection represents the p-values in $-\log_{10}$ scale, and the color of each connection represents the direction of the difference (red: higher in FH- group, blue: higher in FH+ group). The region labels and indices are given in Tab. 4.3

focus our analysis only on these pre-selected COIs. To obtain a large set of initial COIs (to reduce false negatives), we selected the 512 connections in an AD vs. control comparison using our algorithm by applying false discovery rate (FDR) threshold of 0.01. This resulted in a total of 615 COIs out of 13041 connections which were then used to test for effects of family history (FH+ vs. FH-).

Analysis. Seven connections were identified to show significant group differences between FH+ and FH- groups spanning 5 ROIs from the left hemisphere (orbital gray matter, calcarine sulcus, lateral orbital sulcus, postro ventral cingular gyrus and pericallosal sulcus) and 4 ROIs from the right hemisphere (precuneus, superior parietal lobule, posterior sylvian fissure, calcarine sulcus, pericallosal sulcus). These connections are shown in Fig. 4.5 where the color of the connectivity denotes the direction of the strength of connectivity (i.e., red for connections with higher FA in FH- group and blue for connections with higher FA in FH+ group) and the thickness of the connectivity denotes the corresponding p-values (thicker edge for lower p-values). The individual ROIs with white matter pathways connecting the ROIs are visualized in Figs. 4.6 and 4.7. The labels for corresponding pair-wise ROIs with p-values at each connections are listed in Table 4.3 as well.

Table 4.3: Identified connections (between left and right column) and corresponding ROIs showing group differences between FH+ and FH- (controlled for age and sex). Direction -1 / 1 denotes to higher FA value in FH+ / FH- group respectively.

Index	Region Label (Region Index)	Region Label (Region Index)	Direction	p-value
1	ctx_lh_G_orbital (62)	ctx_lh_S_calcarine (102)	-1	2.997e-5
2	ctx_lh_S_calcarine (102)	ctx_lh_S_orbital_lateral (138)	-1	1.218e-4
3	ctx_rh_G_parietal_sup (69)	ctx_rh_G_precuneus (75)	-1	1.889e-4
4	ctx_rh_G_parietal_sup (69)	ctx_rh_Lat_Fis-post (97)	-1	2.86e-4
5	ctx_lh_G_cingul-Post-ventral (34)	ctx_rh_S_pericallosal (147)	1	4.024e-4
6	ctx_lh_S_pericallosal (146)	ctx_rh_S_pericallosal (147)	1	4.396e-4
7	ctx_rh_S_calcarine (103)	ctx_rh_S_pericallosal (147)	1	5.598e-4

Of note are two connections that were observed in both AD vs. control analysis and the FH+ vs. FH- comparison, the connection between left pericallosal (146) and right pericallosal region (147) and the connection between right superior parietal lobule (69) and right precuneus (75). The baseline approach, applying statistical test on the raw FA data, did not reveal any significant connection differences between FH+ and FH-.

4.4.4 Identifying Associations between Longitudinal Connectivity Changes and CSF

4.5 Summary

We presented an algorithm for assessing brain connectivity variations in populations having Alzheimer's disease and in populations that are not diagnosed with AD but have a family history positive risk of AD. The technique, based on performing a Wavelet transform on non-Euclidean spaces such as graphs provides a method for identifying potentially subtle differences between patient groups, and is especially suited to detecting difference due to early, preclinical neurodegeneration. Our study demonstrated extensive connectivity differences between AD patients and controls that were not evident using standard approaches. In addition, we identified connectivity differences due to increased risk for AD, differences that were not observed using a standard approach.

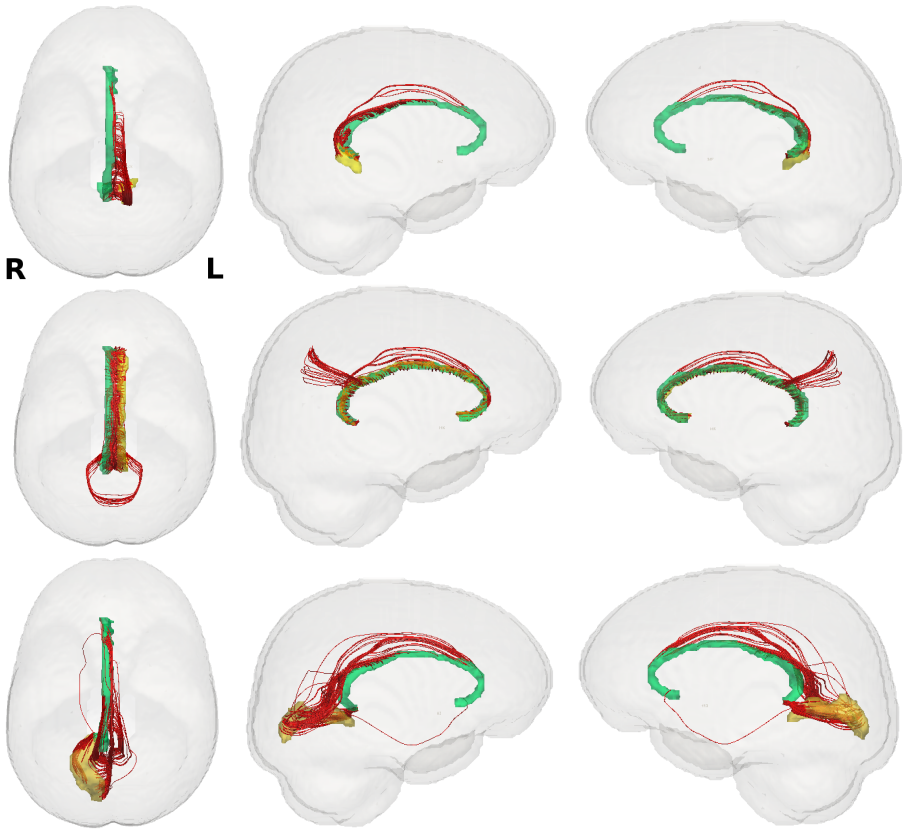


Figure 4.6: Visualization of the white matter tracks of connections (in red) with higher FA values in FH- group. Left column: top view, Middle column: left view, Right column: right view. First row: connection between left ventral posterior cingulate (yellow) and right pericallosal sulcus (green), Second row: connection between left pericallosal sulcus (yellow) and right pericallosal sulcus (green), Third row: connection between right calcarine sulcus (yellow) and right pericallosal sulcus (green). The connection between left pericallosal sulcus (146) and right pericallosal sulcus (147) in the second row is detected in both W-ADRC and WRAP studies.

It is widely accepted that AD has a long preclinical phase during which the brain shows continued degeneration prior to the manifestation of cognitive symptoms. Substantial evidence suggests that the earliest pathology in AD involves abnormal processing of β-amyloid peptide, with the earliest evidence of the disease likely to manifest as alterations to amyloid-related bio-markers (Jack Jr et al., 2010). In addition to amyloid pathology, AD neuropathology is characterized by hyperphosphorylation of tau protein (the protein which

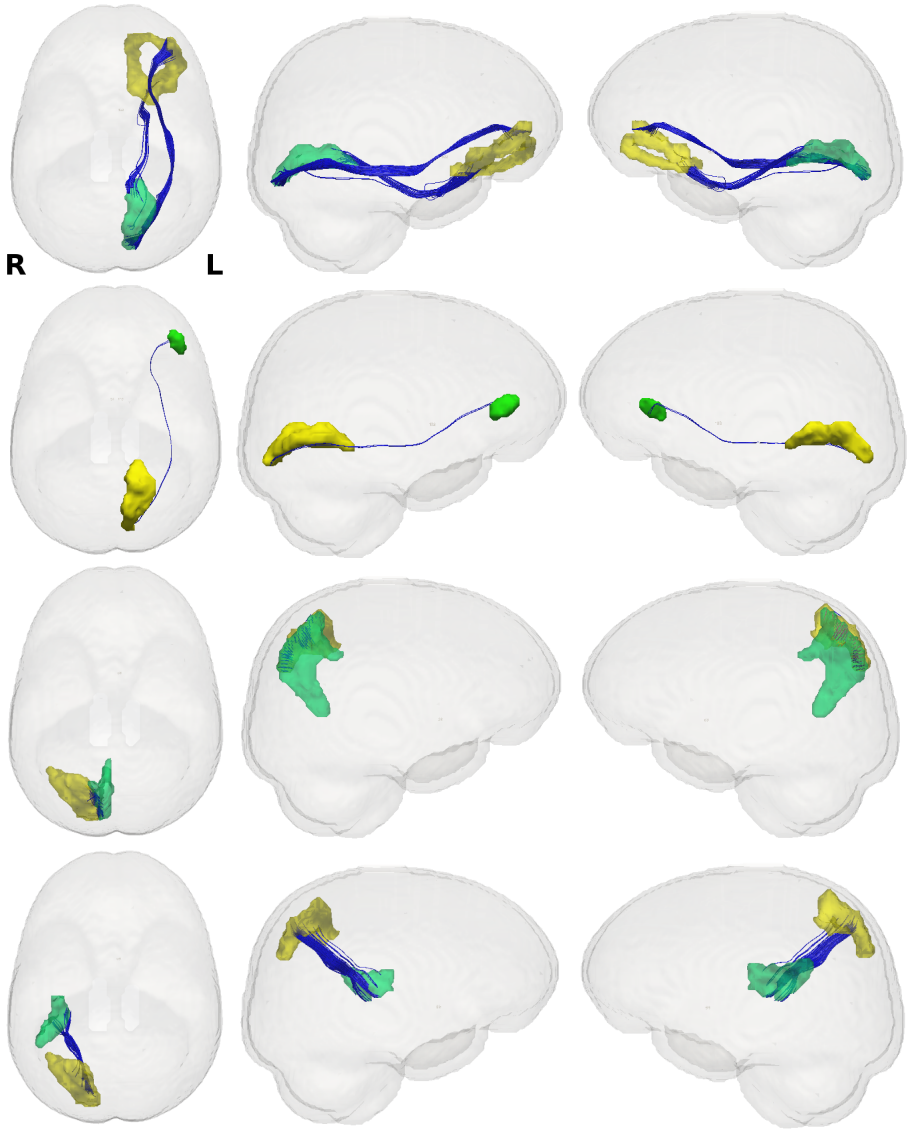


Figure 4.7: Visualization of the white matter tracks of connections (in blue) that have higher FA values in FH+ group. Left column: top view, Middle column: left view, Right column: right view. First row: connection between left orbital gyrus (yellow) and left calcarine sulcus (green), Second row: connection between left calcarine sulcus (yellow) and left lateral orbital sulcus (green), Third row: connection between right superior parietal lobule (yellow) and right precuneus (green), Fourth row: connection between right superior parietal lobule (yellow) and right posterior lateral fissure (green).

stabilizes microtubules in neural cells), and extensive loss of synaptic connections. Increasing evidence suggests that the cognitive dysfunction found in AD may be due to disconnection between highly-interrelated brain regions (Delbeuck et al., 2003; Brier et al., 2014a). Of the characteristics which define AD, neuronal loss and synaptic pathology show the strongest relationship to dementia severity and cognitive deficits in AD (Gómez-Isla et al., 1997; Lassmann et al., 1993; DeKosky and Scheff, 1990; Perez-Nievas et al., 2013).

Among healthy individuals, neural connections underlie normal information processing, and constrain neural activity. Using statistical analysis to identify brain regions which show temporal coherence during fMRI, researchers have identified brain regions that show high-interconnectedness, for example, regions identified using resting-state or task-free fMRI that form the so-called Default Mode Network. Brain regions in this network include the medial prefrontal cortex, medial temporal lobe, and posterior cingulate cortex/retropslenial cortex. These brain regions are part of the episodic memory network (Xu et al., 2009; Johnson et al., 2006), in addition to being active during states of self-awareness (Craik et al., 1999; Gusnard et al., 2001; Fingelkurts et al., 2012; Johnson et al., 2007) and self-monitoring (Schmitz and Johnson, 2007). Interestingly, these brain regions are also those which show some of the highest levels of amyloid burden (Buckner et al., 2005) in AD. Perhaps not surprisingly then, several studies have found that AD patients show altered default mode network (DMN) activity which suggests a loss of connectivity. The finding of altered networks are robust in patients with AD or mild cognitive impairment (MCI) (Sorg et al., 2007; Greicius et al., 2004; Koch et al., 2014). More recently, fMRI based connectivity differences have also been observed in preclinical individuals at risk for the disease (Sperling et al., 2009; Brier et al., 2014b; Sheline et al., 2010).

In order to understand the structural basis for the connectivity loss observed in AD, recent studies have incorporated DTI-based information to determine the extent of loss of myelinated neuronal axons in the disease. In healthy adults, Grecius et al have shown that DMN regions are connected via major white matter tracts (Greicius et al., 2009). While this in itself is not highly surprising, it may suggest that even subtle damage to white matter

tracts could potentially affect network efficacy. In support of this, Hahn et al have shown that disrupted white matter connectivity is associated with altered functional connectivity in MCI and AD (Hahn et al., 2013).

In the present study, we observed that AD affected no less than 81 white matter connections. The majority of the connectivity differences were centered on key brain regions, some of which are in the DMN, and that included portions of the lateral parietal lobe, precuneus, occipito-temporal brain regions, and hippocampus. Our analysis also revealed connectivity differences in tracts connecting to the occipital cortex, which while not a region of high amyloid burden, does show lower cerebral perfusion in MCI (Ding et al., 2014) and altered white matter in MCI patients who convert to AD. In individuals at risk for developing AD, we observed altered connectivity between similar brain regions to those observed in the AD comparison, including precuneus, lateral parietal lobe, and the gray matter of the pericallosal sulcus (which separates the cingulate from the corpus callosum). As with the AD comparison, individuals with family history of AD also showed altered occipital connections. No connectivity differences were detected using standard approaches.

Parental family history of AD is associated with altered connectivity even in asymptomatic adults. Parental family history of AD has in recent years been associated with several brain differences, including altered glucose metabolism (Mosconi et al., 2009), differences in BOLD signal during episodic memory (Xu et al., 2009), lower cerebral perfusion (Okonkwo et al., 2012), increased amyloid deposition, and lower gray matter volume (Mosconi et al., 2014), all in advance of any clinical symptoms of AD dementia. Resting-state connectivity differences have also been observed in individuals who harbor a combination of parental family history risk and APOE ϵ4 carriage. Fleisher et al found differences in connectivity between posterior cingulate/retrosplenial cortex and several cortical regions in the DMN, including higher connectivity with prefrontal and temporal regions in high risk adults, and lower connectivity with precuneus (Fleisher et al., 2009). In addition to altered connectivity between several DMN regions observed in the AD versus control comparison, precuneus emerged as a hub region in the current study, and showed altered connectivity in the FH+ group. The findings also align with prior work from

our group, suggesting that white matter changes are manifested in individuals with parental family history of AD (Adluru et al., 2014; Bendlin et al., 2010), including altered cingulum white matter, which carries fibers interconnecting precuneus (van den Heuvel et al., 2008) with other DMN regions. While the mechanism underlying increased vulnerability to AD based on family history is not known, both genetic and shared environmental factors may play a role (Huang et al., 2004).

It is worth noting that the direction of the connectivity differences between groups was uniform. While the majority of the connections in the AD vs. control comparison revealed higher FA in healthy individuals, both the AD comparison, and the comparison by family history, revealed some connections where FA was higher in the AD, or the at risk group. While higher FA in a disease group is typically unexpected, accumulating studies suggest that selective axonal loss may result in higher FA in the disease group. A study comparing presymptomatic and symptomatic carriers of the presenilin 1 mutation that results in familial AD, has also found that patients in the asymptomatic disease stage exhibit higher regional FA compared to healthy controls(Ryan et al., 2013).

Limitations. A key limitation in this particular application is the physiologic interpretation of higher tract connectivity in AD and FH+. Whether this is due to actual increased connection strength of the tract or is an outgrowth of disease-related simplification / pruning / loss of crossing fibers allowing the tract to be visualized better in the disease group. Separate from this motivating application, we believe that the algorithm can be broadly applied to other types of brain connectivity analyses where the core advantages of higher sensitivity, due to a multi-resolutional perspective, should carry over with very few modifications in an analysis pipeline. Despite the various benefits of the algorithm, there are a few additional limitations that we must point out. First, note that the multi-resolution strategy and our framework involves a quadratic dependence of the number of connections on the number of regions. For example, the number of edges generated in a line graph from a node of degree d will be $O(d^2)$. Even for a fully connected graph, with N number of nodes in the original graph, the number of nodes in the line graph is upper bounded

by the number of edges, $O(N^2)$. In our experiments, each network yields a sparse graph (due to a sparse adjacency matrix), so we did not encounter any scalability issues. Independent of computational issues, when the number of edges is large, the multiple comparisons correction will be fairly strict although our framework does provide means of detecting stronger differences that have a better chance of surviving the correction. A heuristic solution here is to use the multi-resolutional view to come up with a tessellation of the graph into smaller sub-graphs and perform the analysis on sub-graph summaries instead. The specifics will clearly depend on the needs of the application and are not investigated here. Second, as the number of regions p grows, so does the size of the matrix whose decomposition we must obtain. For the sizes used in this work, a Chebyshev polynomial approximation used in the spectral graph wavelet toolbox (Hammond et al., 2011) to approximate the wavelet transformation on graphs is sufficient. Finally, while the method can characterize signals that are scale dependent, when too many scales are defined over the eigenvalue spectrum, the true signal may leak between different scales due to sampling. Choosing special types of wavelet such as "Meyer" wavelet with a tight frame (instead of an overcomplete basis) may help in such cases, but was not needed in our experiments.

Is there a fundamental statistical reason why a multi-resolution view should improve power? Notice that the goal of reducing Type 2 errors is also common in other areas of science (outside neuroimaging) such as statistical genomics. In particular, for analyzing differential gene expression Dahl and Newton proposed a very interesting idea (and accompanying statistical analysis) for improving power in microarray data analysis (Dahl and Newton, 2007). The rationale in that work is elegant yet simple. Consider a setting where the hypothesis tests are being performed on individual genes. Now, if one could define 'true' clusters of genes in terms of shared parameter values, *one could improve the sensitivity of individual gene-level tests, because more data bearing on the same parameter values are available*. In other words, if two genes were part of the same cluster, any degree of information sharing *between them* directly improves the power for hypothesis tests performed for *each of them* because in the most optimistic setting, we have twice as many measurements. The

analysis in their work makes these arguments rigorous and develops ways of obtaining such a clustering based on non-parametric Bayesian methods. In the current work, we do not explicitly 'cluster' the edges of the graph. Nonetheless, interestingly, it is easy to think of a diffusion process on the graph (achieved by wavelets) as, in fact, serving a similar goal as clustering and taking multiple edges into consideration at once. By aggregating information at different/-multiple resolutions (governed by the scaling and dilation parameters of the wavelet expansion), the framework described here offers improved sensitivity based on roughly similar principles. That is, part of the reason why our proposed framework achieves better sensitivity is by considering multiple network edges jointly similar to the more recent work on hyper graph (Davison et al., 2015).

5 STATISTICAL ANALYSIS OF LONGITUDINAL DATA WITH SYSTEMATIC VARIATIONS

As we discussed earlier in Chapter 1, statistical analysis of longitudinal or cross sectional brain imaging data to identify effects of neurodegenerative diseases is a fundamental task in various studies in neuroscience. However, when there are systematic variations in the images due to parameter changes such as changes in the scanner protocol, hardware changes, or when combining data from multi-site studies, the statistical analysis becomes problematic. Motivated by this scenario, the goal of this chapter is to develop a methodological solution to the problem of systematic variations in statistical image analysis. Based in part on recent literature in harmonic analysis on diffusion maps and the ideas in the earlier chapters, we propose an algorithm which compares operators that are resilient to the systematic variations. These operators are derived from the empirical measurements of the image data and provide an efficient surrogate to capturing the actual changes across images. To evaluate the proposed ideas, we later present various experimental results on detecting changes in simulations as well as analysis on real longitudinal PIB-PET imaging data acquired from participants at risk for Alzheimer's disease (AD) to show how the method offers improvement in a statistical analysis.

5.1 Overview

Statistical analysis of a cohort of brain imaging scans, as introduced in Section 1.1, to assess the long term effects of trauma/stress and identify genetic, demographic and lifestyle factors for neurodegenerative diseases is a cornerstone of current research in neuroscience (Canales-Rodríguez et al., 2013; Ruttimann et al., 1998; Chung, 2006). There are two basic but important issues that we can emphasize in such a statistical brain image analysis. First, our ability to conclude that (at a specific voxel) the observed empirical intensity distributions are different across groups depends on the sample size and how distinct the distributions are (i.e., the effect size). Second, this analysis assumes that

the *absolute* image intensity measurements are meaningful. In other words, we assume that the only differences between the groups is due to the effect of the clinical phenomena under study (i.e., age, disease and so on), and *not* other global systematic variations coming from modifications in acquisition parameters. Generally, in small to medium sized studies where the data is acquired at a single site (with the same scanner), this is not a problem. But as scientific studies investigate more subtle scientific questions where the group differences are *weaker*, we need larger sample sizes — logistic constraints necessitate multi-site studies. Changes in the hardware and pulse sequences (and many other factors) across sites introduce systematic variations in the dataset. In fact, even in small studies, a hardware upgrade (between baseline and follow-up acquisitions) may be a nuisance for analysis, requiring ad-hoc normalization which may affect statistical power of detecting true group effects (Driemeyer et al., 2008). When the effect sizes are poor, performing inference on the data without appropriate adjustments could affect the success or failure of the scientific hypothesis under investigation.

The above problem is common across various imaging modalities in medical imaging. For instance, in neuroimaging uses of positron emission tomography (PET), a nuclear imaging modality (where an injected radiotracer binds to specific pathologies as discussed in Section 1.1), image measurements vary considerably, even for the same subject, due to a variety of reasons. So, before any statistical analysis can be performed, these images must be "normalized". Possible approaches include global normalization (mean intensity) or regional scaling (by a reference region). This process converts the intensities into a physiological range of interpretable values. But if the global average or the mean intensity of the reference region used is *not* independent of the condition being studied, the analysis will invariably suffer (an example shown in Fig. 5.1). In these cases, incorrect normalization can lead to an inability to identify real group differences, or worse, one may obtain paradoxical or "opposite" findings. In various other imaging modalities, a normalization strategy may not even be viable. For example, if the systematic variations are the result of changes in the acquisition parameters at different sites, one must analyze the smaller datasets separately. The goal of this chapter is to develop

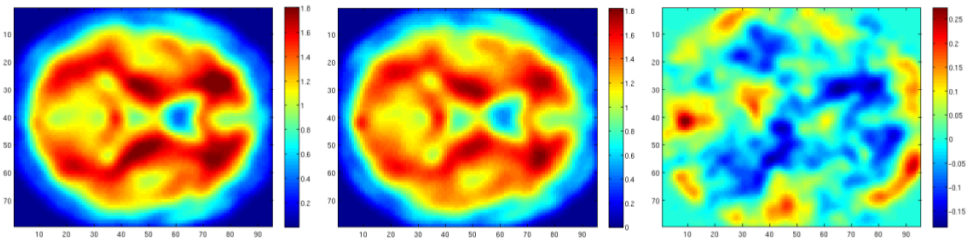


Figure 5.1: An example of systematic variation in longitudinal FDG-PET scans. Left: FDG-PET scan of a subject, Middle: FDG-PET scan of the same subject after 2 years, Right: the changes between the two scans. FDG intensities tend to increase over time with glucose uptake, but significant decreases are shown in many regions.

a unified statistical solution to this problem.

5.1.1 Related Work

There are several broadly related ideas in vision and medical imaging that can serve as a reasonable starting point for comparing functions that cannot otherwise be compared (Eismann et al., 2008). The most natural choice is a statistical measure that is, by construction, invariant to image intensities: Mutual Information (MI). Mutual information has been extensively used in both computer vision (e.g., stereo (Hirschmuller, 2005; Kim et al., 2003; Hirschmuller, 2008)) and in medical imaging (e.g., non-linear registration (Viola and Wells III, 1997; Maes et al., 1997; Pluim et al., 2003; Wells III et al., 1996)) and offers precisely the type of invariance we desire. While MI is a good loss function to optimize when searching for a non-linear transformation or disparity map, once such a transformation has been found and the images have been aligned, MI does not make the statistical analysis any easier. For instance, consider a set of ten participants whose images were acquired twice, a few years apart, and the intensities in the second acquisition are systematically different (e.g., affine scaling). While MI *can* characterize the joint entropy of a pair of intensities, it *cannot* be easily used to quantify the voxel-wise change from one time point to the other.

An alternative to the MI approach is based on dictionary learning/patch regression inspired idea called *image synthesis* (Iglesias et al., 2013; Roy et al.,

2010; Osman and Prince, 2004). Broadly, one may use image synthesis to synthetically generate the image that has been corrupted, assuming a large set of training examples is available. While this approach is suitable for addressing missing data, applying it in the above longitudinal setting will entail generating the *entire* set of images at the second time point. The learning task will broadly correspond to inferring the parameters of a generative model that explains *temporal change* across the population, given only the baseline acquisition — clearly difficult regardless of how well characterized the training dataset is. Given these issues, to our knowledge, there is no universally applicable solution offering the same capabilities as the algorithm we propose here. In situations where the structural variations in the intensities are related by a simple transformation, one may normalize the entire image by a suitable normalization constant. In medical imaging, this is often difficult because it must be derived from a region not affected by disease, age, or the clinical phenomena under study. If this is sub-optimal, it can affect the statistical analysis in unexpected ways.

5.1.2 A High Level Description of the Idea

Let f denote an unknown function. Let α and β denote two parameters such that they modify the form of the function $f(\cdot)$ yielding f_α and f_β. Now, consider that we are only given access to measurements of f_α and f_β. It is clearly not possible to verify whether they were both derived from the same latent function f, unless we also know the relationship between the transformations of f induced by α and β (if the respective inverse transformations are unique). Assume that an oracle provides us an operator $\mathcal{T}$ with the interesting property that it is *invariant* to the parameter space $\mathcal{P}$ from which α and β are drawn. That is, if we construct a pair of operators from the empirical measurements of f_α and f_β, the operators will be the same: $\mathcal{T}_{f_\alpha} \equiv \mathcal{T}_{f_\beta}$ if they share the same latent function, f.

Next, consider a slightly more complicated setup. The latent function f has now been modified to f'. We are now provided with the measurements, f_α and f'_β, i.e., both the parameter and the function change. Since the opera-

tor $\mathcal{T}$ only offers invariance to the parameter space $\mathcal{P}$ (and assumes that the latent function is the same), in this case, the operators $\mathcal{T}_{f_\alpha}$ and $\mathcal{T}_{f'_\beta}$ *cannot* be compared. Nonetheless, we can see that the operators provide a mapping to two *different* spaces, say S_{f_α} and $S_{f'_\beta}$, since f and f' are distinct. Interestingly, because of the invariance to $\mathcal{P}$, if we now plug in a *known* function (such as an impulse function) at all locations in the original space into the two operators, we will obtain its transformed representations in S_{f_α} and $S_{f'_\beta}$. Once these transformed forms of the impulse functions are mapped from $S_{f'_\beta}$ to S_{f_α}, we can calculate the distance. If the distance is near zero, then $f \simeq f'$; otherwise, it characterizes the discrepancy between f and f' since the operators are, by design, invariant to $\mathcal{P}$. In the following sections, we demonstrate how this idea can be implemented using wavelet expansion in non-Euclidean space described in Chapter 2.

5.2 Wavelet Map and Wavelet Kernel Distance

Defining a kernel function in a square integrable measure space (X, μ) enables one to measure local similarities within X at small scales (Coifman and Hirn, 2014). We therefore define a mother wavelet function as such a kernel function using an operator $\mathcal{T}^s$, which is constructed using empirical measurements of function f_α. A mother wavelet $\psi_{s,p}(q)$ described in Section 2.2 can be viewed as if it were a kernel function written as $\psi_s(p, q)$, defining a relationship between vertex p and q (Brislawn, 1991). Using this, we define Wavelet Kernel Distance (WKD) $d_s(p, q)$ at scale s, a measure between two points p and q defined as ℓ^2—norm of the wavelet density difference over the space X as

$$d_s(p, q)^2 = \|\psi_s(p, \cdot) - \psi_s(q, \cdot)\|_2^2 \tag{5.1}$$

$$= \int_X (\psi_s(p, r) - \psi_s(q, r))^2 \mu(r) \tag{5.2}$$

In the graph setting, using the SGWT operator $\mathcal{T}_g^s$ as in Section 2.2, observe that (5.2) can be rewritten using a wavelet kernel function $g()$ in the spectral

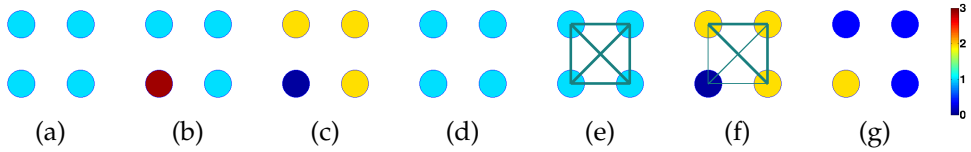


Figure 5.2: Comparing two differently scaled functions on four data points. a) $f(p) = f_\alpha(p) = (1,1,1,1)^T$, b) $f'(p) = (1,1,3,1)^T$, c) $f'_\beta(p) = 3 - f'(p) = (2,2,0,2)^T$, d) $|f_\alpha - f'_\beta|$, e), f) graphs from f_α and f'_β respectively (edge thickness denotes to edge weight), g) WKD using structure from e) and f). The true change between a) and b) is $(0,0,2,0)$, but a simple subtraction in d) is inaccurate. The proposed algorithm can capture the true change in g).

domain as

$$d_s(p,q)^2 = \sum_{l=0}^{N-1} g(s\lambda_l)^2(\chi_l(p) - \chi_l(q))^2 \tag{5.3}$$

It can be interpreted as if we were comparing the effect of the same wavelet function dissipating from different locations p and q to their neighbors by the wavelet kernel function $g()$, thereby measuring the effect of the propagation.

Further, we can also define a mapping of δ_p at each grid-point to a lower dimensional Euclidean space spanned by χ defined as the wavelet map $\gamma : X \to \ell^2$ at scale s as

$$\gamma_s(p) = (g(s\lambda_l)\chi_l(p))_{l=0,1,\ldots,N-1} \tag{5.4}$$

characterizing the local relationship of the graph with the wavelet kernel function $g()$. Note that when $g_s(\lambda_l) = \lambda_l^s$, the wavelet map exactly becomes diffusion map proposed earlier in (Coifman and Lafon, 2006).

A toy example is shown in Fig. 5.2: the objective here is to compare two different functions f_α and f'_β defined on four data points, and find the true difference between them. Given latent functions $f = (1,1,1,1)^T$ and $f' = (1,1,3,1)^T$, the true difference (i.e., $|f - f'|$) here is $(0,0,2,0)$. Given the latent functions, f_α remains the same as f while f'_β is defined to be $f'_\beta(p) = 3 - f'(p)$. Clearly, a direct comparison of f_α and f'_β (i.e., $|f_\alpha - f'_\beta|$), as illustrated in Fig. 5.2 d), fails to detect the true difference. On the other hand, computing WKD from graphs constructed using f_α and f'_β at each data point yields the true difference as shown in Fig. 5.2 g).

We can now formally establish the relationship between wavelet map, WKD, and the construction of Wavelets using the following two results.

Proposition 5.1. *The squared WKD* d_s *defined between two vertices* p *and* q *on the same graph is equivalent to the* ℓ^2*—norm of the difference between the respective wavelet maps of vertices* p *and* q.

Proof. Taking the ℓ^2—norm of the difference over wavelet map on vertices p and q yields,

$$d_s(p,q)^2 = \|\gamma_s(p) - \gamma_s(q)\|_2^2 \tag{5.5}$$

$$= \sum_{l=0}^{N-1} (g(s\lambda_l)\chi_l(p) - g(s\lambda_l)\chi_l(q))^2 \tag{5.6}$$

$$= \sum_{l=0}^{N-1} g(s\lambda_l)^2(\chi_l(p) - \chi_l(q))^2. \qquad \square$$

From Proposition 5.1, we can see that WKD defines a Euclidean distance of the wavelet maps between vertices p and q in the space formed by χ. We can further define a relationship between Wavelet maps and an actual wavelet function.

Proposition 5.2. *Let* $\mathcal{V} = [\chi_0\ \chi_1\ \dots\ \chi_{N-1}]$ *denote a matrix where* χ_i *corresponds to columns. The projection of a wavelet map* $\gamma_s(p)$ *at vertex* p *to the row space of* $\mathcal{V}$ *precisely constructs a mother wavelet function* $\psi_{s,p}(q)$.

Proof. Given $\chi(q)$, the qth row of $\mathcal{V}$, taking inner product of the wavelet maps $\gamma_s(p)$ and $\chi(q)$ becomes

$$\langle \gamma_s(p),\ \chi(q) \rangle = \sum_{l=0}^{N-1} g(s\lambda_l)\chi_l(p)\chi_l(q) \tag{5.7}$$

$$= \psi_{s,p}(q) \quad (\text{since } \chi_l^* = \chi_l) \tag{5.8}$$

which defines a mother wavelet function at q centered at p exactly in the form given in (2.10) . $\square$

We see that Proposition 5.2 establishes the connection between the construction of wavelet basis from Section 2.2 and the wavelet map. It shows that a Wavelet function can be constructed from the wavelet map at each vertex. Further, this result ties the wavelet map to kernel signatures on graphs, variants of which have been used for graph matching and surface segmentation (but using diffusion (Fang et al., 2011; Sun et al., 2012)). When the wavelet map of p is projected to the pth row of $\mathcal{V}$, we get a wave-type kernel descriptor in (Aubry et al., 2011; Kim et al., 2013b). Separately, when $g_s(\lambda_l) = \exp(-s\lambda_l)$, we obtain the heat kernel signature in (Bronstein and Kokkinos, 2010).

5.3 Generalization of Wavelet Kernel Distance

So far, we have shown how two different vertices on the same image/graph can be compared using a Wavelet operator $\mathcal{T}_{f_\alpha}$ that has been derived from empirical measurements of a function f_α. We further generalize the method to compare two different vertices from two different graphs by looking into two different operators.

Consider two individual graphs I and J, constructed using functions (or images) f_α and f'_β, where the number of vertices in each is N. We assume that the vertices are spatially registered and that we are operating on a square integrable space X. On these graphs, WKD between a vertex p^I from I and a vertex q^J on J is defined as

$$d_s(p^I, q^J)^2 = \|\psi_s^I(p,\cdot) - \psi_s^J(q,\cdot)\|_2^2 \tag{5.9}$$

$$= \int_X (\psi_s^I(p,r) - \psi_s^J(q,r))^2 \mu(r) \tag{5.10}$$

using wavelet kernel functions ψ_s^I and ψ_s^J.

We first construct two operators $\mathcal{T}_{f_\alpha}$ and $\mathcal{T}_{f'_\beta}$, and obtain two sets of orthogonal bases χ^I and χ^J from each operators to compare the vertex-wise differences. Note that while the expansion of (5.9) does not simplify as in (5.3) since the eigenvectors χ^I and χ^J are no longer orthogonal to each other, it nonetheless reduces to a meaningful expression defining a mapping between

the lower dimensional spaces defined by the two operators as described by the following result.

Proposition 5.3. *Let* λ^I, λ^J *and* χ^I *and* χ^J *denote the eigenvalues and eigenvectors from graphs of* I *and* J *respectively. Then, the WKD* $d_s(p^I, q^J)$ *can be written as,*

$$d_s(p^I, q^J)^2 = \sum_{l_1=0}^{N-1} g(s\lambda_{l_1}^I)^2 \chi_{l_1}^I(p)^2 + \sum_{l_2=0}^{N-1} g(s\lambda_{l_2}^J)^2 \chi_{l_2}^J(q)^2 - 2 \sum_{l_1,l_2=0}^{N-1} g(s\lambda_{l_1}^I)\chi_{l_1}^I(p) g(s\lambda_{l_2}^J)\chi_{l_2}^J(q)\langle \chi_{l_1}^I, \chi_{l_2}^J \rangle \tag{5.11}$$

Proof. Let X be a square integrable space and $\{\lambda_l^I, \chi_l^I\}$ and $\{\lambda_l^J, \chi_l^J\}$ be eigenvalue and eigenvector pairs from graph Laplacians of graph I and J respectively.

$$\begin{aligned} d_s(p^I, q^J)^2 &= \|\psi_s^I(p, r) - \psi_s^J(q, r)\|_2^2 && (5.12)\\ &= \int_X (\psi_s^I(p, r) - \psi_s^J(q, r))^2 d\mu(r) && (5.13)\\ &= \int_X \psi_s^I(p, r)^2 + \psi_s^J(p, r)^2 - 2\psi_s^I(p, r)\psi_s^J(q, r) d\mu(r) && (5.14)\\ &= \sum_{l_1=0}^{N-1} g(s\lambda_{l_1}^I)^2 \chi_{l_1}^I(p)^2 + \sum_{l_2=0}^{N-1} g(s\lambda_{l_2}^J)^2 \chi_{l_2}^J(q)^2 && (5.15)\\ &\quad - 2 \sum_{l_1,l_2=0}^{N-1} g(s\lambda_{l_1}^I)\chi_{l_1}^I(p) g(s\lambda_{l_2}^J)\chi_{l_2}^J(q)\langle \chi_{l_1}^I, \chi_{l_2}^J \rangle \end{aligned}$$

□

Let us look at the individual terms in (5.11) more carefully. The first two terms in (5.11) form the WKD on a single graph whereas the last term compensates for the discrepancy caused by the variations of the inherited spaces once the first space has been mapped to the other. By inspection, we see that this generalizes Proposition 5.1. When I and J are the same, we can easily verify that,

Proposition 5.4. *When* I *and* J *are equal, then* (5.11) *reduces to* (5.3).

Proof. Since I and J are the same graph, they share the eigenvalues λ_l and eigenvectors χ_l, therefore

$$\begin{aligned} d_s(p^I, q^J)^2 &= \sum_{l_1=0}^{N-1} g(s\lambda_{l_1}^I)^2 \chi_{l_1}^I(p)^2 + \sum_{l_2=0}^{N-1} g(s\lambda_{l_2}^J)^2 \chi_{l_2}^J(q)^2 \\ &\quad - 2 \sum_{l_1,l_2=0}^{N-1} g(s\lambda_{l_1}^I)\chi_{l_1}^I(p) g(s\lambda_{l_2}^J)\chi_{l_2}^J(q)\langle \chi_{l_1}^I, \chi_{l_2}^J \rangle \\ &= \sum_{l=0}^{N-1} g(s\lambda_l)^2 (\chi_l(p)^2 - \chi_l(q)^2) \end{aligned}$$

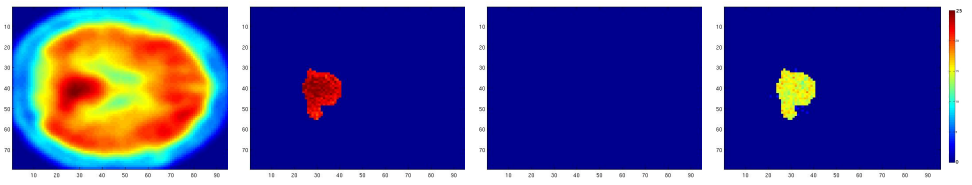
with $\langle \chi_{l_1}^I, \chi_{l_2}^J \rangle = 0$ when $l_1 \neq l_2$. □

5.4 Group Analysis of Synthetic PIB images

We now present results of statistical analysis on a population of synthetically generated 2-D Pittsburgh Compound B (PIB) image data. The experiment design is as follows. We assume we have two groups: diseased and healthy (controls). We simulate brain images of 20 diseased and 20 control subjects, using a template 2-D PIB image with size of 79×95. We assume that each subject was imaged longitudinally providing a t_0 (baseline) and t_1 (follow-up) image. At t_0, the images Y_{t_0} in both (diseased and control) groups are modeled as a random field with mean μ_{control} with added Gaussian noise $N(0, 0.1)$ as

$$Y_{t_0} = \mu_{\text{control}} + N(0, 0.1) \tag{5.16}$$

where μ_{control} is given by the template PIB image slice shown in Fig 5.3 (a). At t_1, we consider two types of changes: the first is an increase of PIB values by 20% in certain regions of the brain in the diseased group characterized by μ_{disease}, and the other is systematic variation simulated as an arbitrary affine transformation with scale $s \in [1, 2]$ and translation $a \in [0, 1]$ applied to the



(a) (b) (c) (d)

Figure 5.3: Result from a group analysis on diseased vs. normal groups using synthetic PIB images. a) a template PIB image used for the mean μ, b) p-value map in $-\log_{10}$ scale from the group analysis using images without the systematic variation (serving as the ground truth), c) p-value map in $-\log_{10}$ scale from the group analysis using images with systematic variation, d) p-value map in $-\log_{10}$ scale from the group analysis using WKD on images with systematic variation. We can see that using WKD, we can detect group differences even when there is a systematic variations in the images.

image intensities.

$$Y_{t_1} = \begin{cases} s\mu_{\text{control}} + a + N(0, 0.1) & \text{if normal} \\ s\mu_{\text{disease}} + a + N(0, 0.1) & \text{if diseased} \end{cases}$$

In this scenario, we would like to detect the changes $\Delta Y = Y_{t_1} - Y_{t_0}$ from the two time points across the two groups by comparing the distribution of ΔY across groups. In the standard procedure, performing a statistical hypothesis test at each pixel (a total of 7505 tests) yields a p-value at each pixel, that tells us whether the distribution of the ΔY are the same. Applying Bonferroni correction at 0.05 removes false positives and identifies the regions with significant changes between the two groups. This process works well when $s = 0$ and $a = 0$, however, systematic variations may reduce or bias the effect sizes and diminish the statistical power. Using our method, we expect to detect the group differences even in the presence of systematic variation.

The resultant p-value maps from this simulation is displayed in Fig. 5.3 b), c) and d) at the same scale ($-\log_{10}$ scale), which shows three cases of this experiment: using the standard hypothesis testing procedure on (i) the given data without systematic variations (i.e., ground truth), (ii) with systematic variations and (iii) WKD for the data with systematic variations. As seen in Fig. 5.3 (b), there is a strong signal showing group differences between the two groups (diseased and controls), easily identified using standard hypothesis

testing. This serves as the ground truth. In contrast, when there are systematic variations in the data, the traditional approach fails to detect the true differential signal as shown in Fig. 5.3 c). We computed WKD at each pixel of the images with systematic variations instead of computing ΔY directly, and then applied hypothesis testing on WKD. This process successfully detects the region as shown in Fig. 5.3 d) showing excellent consistency with the actual changes between t_0 and t_1. Therefore, in this sanity check experiment, our method correctly picks up the true variations and makes the downstream statistical analysis more sensitive even when systematic variations exist.

5.5 Analysis of Longitudinal PIB Changes

Here, we demonstrate results from a longitudinal PIB-PET image analysis, where we use the ratio of total τ protein and amyloid-β-142 (Aβ(1-42)) as a predictor for the increase in voxel-wise PIB values at two different time points. PIB values are used as a measure of brain amyloid deposition, a core pathological feature of Alzheimer's disease (AD), and it is known that such increase is closely correlated with AD. The Aβ(1-42) interacts with the signaling pathways to control the phosphorylation of τ protein (LaFerla, 2008; Ittner and Götz, 2010) and their ratio is widely used as a sensitive feature of AD pathology (Fagan et al., 2007; Diniz et al., 2008).

5.5.1 Dataset

The dataset of 84 participants from WRAP study, including subjects that are otherwise healthy but may have potential risk factors for AD is used for the experiment. The participants are comprised of 26 males and 58 females with the mean age 67.4, and the 3-D PIB images from each participants are spatially registered to the Montreal Neurological Institute (MNI) space. The image intensities represent standard uptake value (SUV), which is the ratio of the tissue radioactivity concentration and injection divided by the body weight. These values are scaled with the intensity from a reference region (i.e., cerebellum), generating standard uptake value ratio (SUVR) images (Lucignani et al.,

2004). The PIB intensities, by nature, tends to increase due to accumulation of amyloid burden in the brain. However, when the SUVR images between two time points are compared (i.e., affected by time), various brain regions show decrease in the PIB values, indicating that there exist systematic variations that have not been account for by the normalization process.

5.5.2 Experimental Setup

For the graph representation of each volume image, we used a grid graph with six neighbors for each voxel in 3-D space. The connection between voxels were defined by $\exp(-\|I(p) - I(q)\|^2/\sigma^2)$ where $I(p)$ is the PIB intensity at voxel p and $\sigma = 0.1$, and used a Jacobi-Davison conjugate gradient method (Notay, 2002) to compute the first fifty eigenvalue/eigenvector pairs of the matrix. For the wavelet kernel function g, we used the cubic spline function provided in SGWT (Hammond et al., 2011).

5.5.3 Experimental Results

A high positive correlation between the PIB changes and the ratio between total τ protein and Aβ(1-42) indicates that the increase of the PIB values are highly related to the increase of the ratio. When compared to the result using SUVR images, stronger correlation from WKD is demonstrated in larger regions of the brain. Among the total of 510340 voxels, WKD identifies 21101 voxels (4.13%) with correlations above 0.3 — a common threshold for moderate correlation. On the other hand, using SUVR images, we find only 14655 voxels (2.87%) with the same threshold.

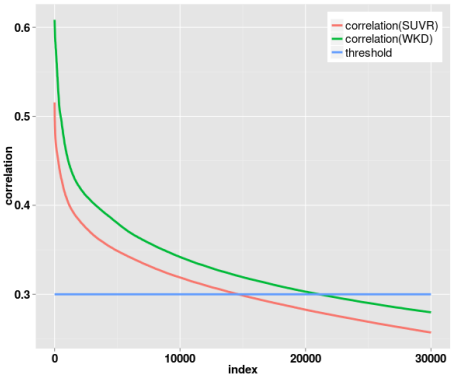


Figure 5.4: Plot of sorted correlation with respect to the number of voxels. The correlation using WKD (green) and SUVR images (red) show that WKD shows higher correlation with larger number of vertices above the threshold (blue).

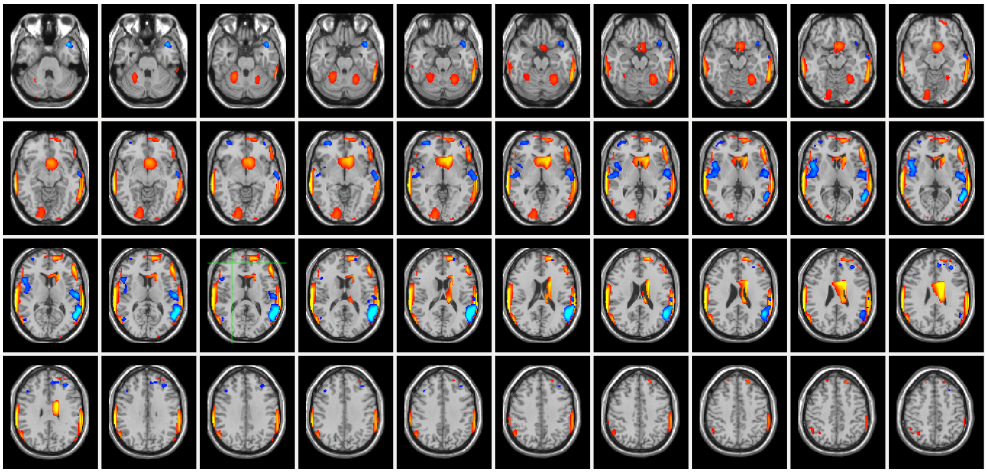


Figure 5.5: Montage of axial view of the correlation between the PIB changes and the ratio of total τ-protein and Aβ(1-42) on a template T1-weighted brain image. The red-yellow intensities indicate correlation using WKD, and the blue-light blue intensities indicates correlation using SUVR images in the range of [0.3 0.5].

The sorted correlations are shown in Fig. 5.4, indicating that WKD is more sensitive than the differences found via SUVR images.

Fig. 5.5 shows the resultant correlation overlayed on a T1-weighted template, where the correlations using WKD and SUVR images are shown in red-yellow and blue-light blue maps in the same range respectively. The result shows that both our analysis and the one performed on SUVR images agree on moderate correlations in lateral temporal lobe regions, which are well-known to be affected by AD (Chan et al., 2001; Scheltens et al., 1992; Kim et al., 2012, 2014) — but our algorithm shows higher correlation and larger regions. Interestingly, WKD framework also picks up the bilateral cerebellum regions which is known to show loss of volume with dementia (Reiman et al., 2009). Note that this region is very close to regions that are used as the 'reference' for the SUVR normalization — therefore will not be identified in the standard analysis even if affected by disease.

5.6 Summary

In this chapter, we proposed a solution to a problem where statistical analysis of imaging data in brain imaging studies is problematic due to systematic variations caused due to a variety of factors. Motivated from recent literature in harmonic analysis, we proposed to compare operators as a means of detecting changes across images, when the absolute measurements cannot be compared on their own. These operators are derived from empirical measurements of images and provide invariance to the systematic variations. Using our framework, we showed experiments on synthetic as well as real datasets, demonstrating that the algorithm works well in a regime where few alternatives are currently available. In particular, in an interesting application to brain imaging data from subjects at risk for Alzheimer's disease, we show that the sensitivity and power of statistical analysis of PIB-PET images can be improved by using the proposed method.

Despite the advantages of the proposed method, there are some potential limitations of the method from the neuroscience point of view. For instance, one issue is that the analysis may miss out on some regions that are found by the standard analysis. In these situations, it is difficult to assess whether this is an artifact of our method or a consequence of the normalization process in the standard analysis.

6 DEFORMATION INVARIANT REPRESENTATION OF IMAGES VIA SCATTERING TRANSFORM ON GRAPHS

A variety of studies in neuroscience/neuroimaging seek to perform statistical inference on the acquired brain image scans for diagnosis as well as understanding the pathological manifestation of diseases. To do so, an important first step is to register (or co-register) all of the image data into a common coordinate system. This permits meaningful comparison of the intensities at each voxel across groups (e.g., diseased versus healthy) to evaluate the effects of the disease and/or use machine learning algorithms in a subsequent step. But errors in the underlying registration make this problematic, they either decrease the statistical power or make the follow-up inference tasks less effective/accurate. In this chapter, we derive a novel algorithm which offers immunity to local errors in the underlying deformation field obtained from registration procedures. By deriving a deformation invariant representation of the image, the downstream analysis can be made more robust as if one had access to a (hypothetical) far superior registration procedure. Our algorithm is based on recent work on scattering transform. Using this as a starting point, we show how results from harmonic analysis on graphs (introduced in Chapter 2) yields strategies for designing deformation and additive noise invariant representations of large 3-D brain image volumes. We present a set of results on synthetic and real brain images where we achieve robust statistical analysis even in the presence of substantial deformation errors; here, standard analysis procedures significantly under-perform and fail to identify the true signal.

6.1 Overview

In most of statistical neuroimage analyses, a key component which precedes much of the "analysis" pipeline is co-registration (Ashburner and Friston, 1997) as we discussed earlier in Chapter 1. For instance, if the study cohort

includes a set of 100 participants/subjects, the co-registration step will *warp* each of the 100 images into a common *template* coordinate system. It is easy to see that doing so enables one to retrieve the measurement at a specific voxel v *across* the entire cohort — in other words, for any arbitrary voxel in one image, we know its corresponding voxels in all other images (demonstrated in the middle of Fig. 1.3). This process, often called as "spatial normalization", is essential for performing either voxel-wise inference (i.e., independently, one voxel at a time) or utilizing such data for classification or regression tasks (Shattuck et al., 2001; Smith et al., 2004; Zhang et al., 2011).

Given the critical role of registration in statistical image analysis of brain imaging data, the community has invested much effort in algorithms and (mature) software libraries that are reliable and easily deployable. Depending on the clinical question and the type of image modality under study, methods based on non-linear/free-form deformations optimizing various loss functions (e.g., mutual information) are broadly available (Maes et al., 1997; Maintz and Viergever, 1998; Roche et al., 1998; Mattes et al., 2003). Note that while the general ideas driving image registration do not change very frequently, most popular toolboxes are regularly fine-tuned, to achieve better accuracy. These "incremental" refinements seem routine, but are very valuable in practice. To see why, let us use an example to answer a closely related question, namely, what happens when the registration pre-processing is slightly imperfect?

Consider a dataset which consists of two groups: healthy and diseased subjects. Once the initial co-registration is done, our goal may be to perform a voxel-wise parametric test to identify voxels that are statistically different across the groups suggesting a potential disease effect (Ashburner and Friston, 2000). It is commonly observed by practitioners that using a newer registration method that is marginally better relative to its older version improves the final analysis in measurable ways: for example, by revealing a *stronger* disease effect (i.e., *lower* p-values) and possibly more image regions where differences are statistically significant (survive a threshold of $\alpha = 0.05$) across the healthy and diseased groups. This suggests that even small errors in the registration may have a detrimental effect on the downstream analysis. The statistical power improvements described above have important and real consequences; an

actual experiment or study that may have needed 500 participants to evaluate a hypothesis may now be possible (using a better registration procedure) with a slightly smaller sample size.

Clearly, image registration algorithms will continue to improve. But separate from these developments, our work here considers a complementary (but potentially more interesting) question that is *not tied to* which method is currently the best. Assume that there is an ideal registration (or transformation), $\mathcal{T}$ and the one estimated by a state of the art registration method is $\mathcal{T} + \epsilon$ with error $\epsilon > 0$. Based on the foregoing discussion, actively pursuing better registration schemes is obviously important. But independent of these improvements, can we derive statistical analysis methods that are, by design, *immune* to nominal values of ϵ? In other words, as long as the registration procedure provides a reasonable estimate of $\mathcal{T}$, the follow-up analysis operates on alternate representations of the image that are *invariant* to local deformations (of up to magnitude $\epsilon' \leqslant \epsilon$) (Kowalski, 2014; Ling and Jacobs, 2005; Morel and Yu, 2009; Lawrence et al., 1997; Simpson et al., 2013, 2011). Such a tool, if available, will have two direct implications. Notice that if the upstream registration is already good, such a local-deformation invariant analysis framework may occasionally offer a small improvement or at worst, will match the results that we obtained anyway. But more importantly, *if* the results from state-of-art registration methods are imperfect, such a framework will be extremely useful. In some sense, the final analysis results will be consistent with a (hypothetical) registration procedure that did *not* make those errors. Note that while spatial smoothing may provide some resilience to such local registration errors, it will be desirable to obtain algorithms that are invariant to such errors. Later in our experimental section, we demonstrate that such a framework yields improved result on a statistical analysis for imperfectly registered (by adding noise to the warping field) Fludeoxyglucose (FDG) PET images.

6.2 Scattering Transform

The scattering transform is known to provide a representation robust to deformation groups. It make use of the wavelet transform and the modulus operator

together to obtain such group invariance. From Chapter 2, the mother wavelet function ψ is defined as a localized oscillating function with finite support. It is this localization property that offers partial robustness to deformation within the classical wavelet transform. Unfortunately, it is again this localization property that makes the wavelet transform *variant* to local translations. To overcome this, the scattering transform uses an additional non-linear operation to make the resultant representations invariant to such localization.

Among the several non-linear operations one can adopt, scattering transform uses the modulus operator, M, which is a piece-wise linear operator. Moreover, the M satisfies the following three important properties: it is 1) point-wise, 2) non-expansive (i.e., $\|Mf - Mg\| \leqslant \|f - g\|$), and 3) signal energy preserving (Bruna and Mallat, 2013).

Having the wavelet transform and modulus operation, the scattering transform obtains its non-linear invariants by combining wavelets (for additive noise invariance), modulus (for local translation invariance) and averaging operation (for local rotation or some other group invariance) in a cascade fashion, and therefore obtains a local deformation invariance. Let $U[s]f = |f \star \psi_s|$, and any sequence $p = (s_1, s_2, \cdots s_m)$ define a path of an ordered product of operators as

$$U[p]f = U[s_m] \cdots U[s_2]U[s_1]f \tag{6.1}$$

$$= |\cdots \| f \star \psi_{s_1}| \star \psi_{s_2}| \cdots \star \psi_{s_1}| \tag{6.2}$$

Using this concept, a scattering transform (represented as a convolution network) along path p at the location u is defined as

$$S[p]f = U[p]x \star \phi(u) \tag{6.3}$$

$$= |\cdots \| f \star \psi_{s_1}| \star \psi_{s_2}| \cdots \star \psi_{s_1}| \star \phi(u) \tag{6.4}$$

yielding scattering coefficients $S[p]f$ for path p. These scattering coefficients are invariant to translation of f and although it has many similarities with the Fourier transform modulus, $S[p]f$ is Lipschitz continuous to deformation in contrast to the Fourier transform modulus. Such a transform can represented

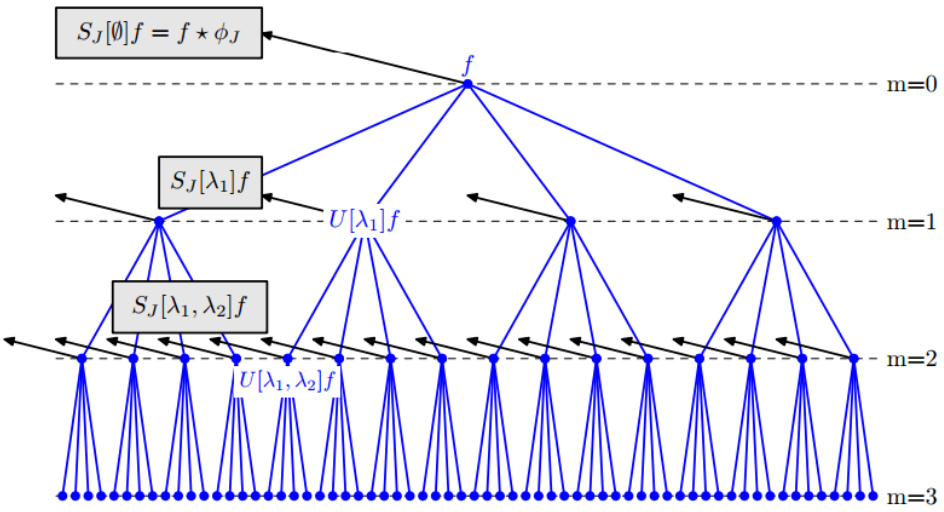


Figure 6.1: Example of scattering transform as a convolution network. Figure adopted from (Bruna and Mallat, 2013).

as a convolution network as demonstrated in Fig. 6.1.

The trouble with above construction of scattering transform is that it is inapplicable where the dimension of the function space and size of the data is large, since the number of rotation parameter increases when the dimension of the function space increases. A potential solution to the issue here is to utilize the underlying structure of the given data using graph and avoid a computational cost that increases rapidly with the dimension. When images are expressed as a graph, the spatial correlations are encoded as edges whereas the number of nodes only depends on the size of the image, not on its dimensionality. In this scenario, intuitively, the notion of rotation can be substituted by strategies based on anisotropic filtering governed by the edge weights. If some additional technical conditions hold, one can still offer the types of properties obtained in the conventional scattering transform.

Recall that wavelet transform is the core of the scattering transform, which classically has been defined *only* in the Euclidean space. In order to define scattering transform on data that lives in the non-Euclidean space (i.e., a graph), we utilize spectral graph wavelet transform from Section 2.2.

6.3 Scattering Transform on Graphs

With the above components in hand, deriving a Scattering transform on a graph can be performed mechanically. First, we define a single scattering operation by combining the wavelet and modulus operations together as,

$$S(t, n) = |f \star \bar{\psi}_{t,n}| \tag{6.5}$$

which yields the local deformation invariant coefficients $S(t, n)$ in scale t at vertex n. Note that in order for a transform to achieve robustness to deformation, *it must be non-expansive*, so we should expect the graph scattering transform presented above to offer these properties. First, $S(\cdot, \cdot)$ is local deformation invariant since $\bar{\psi}$ is one, and it is also invariant to additive noise which is an important property that does not hold for the classical Fourier transform. This can be shown by proving that it is a nonexpansive operator (Mallat, 2012),

Proposition 6.1. *The scattering operation S on a graph using SGWT operator T_g^s at scale s with a kernel $g()$ that is $\int g(x)dx = 1$ is non-expansive.*

Proof. Given two functions f and h defined on a graph,

$$\begin{aligned}
\|Sf - Sh\| &= \big\||T_g^s f| - |T_g^s h|\big\| \\
&= \left\| |\sum_{l=0}^{N-1} g(t\lambda_l)\hat{f}(l)\chi_l| - |\sum_{l=0}^{N-1} g(t\lambda_l)\hat{h}(l)\chi_l| \right\| \\
&\leqslant \left\| \sum_{l=0}^{N-1} g(t\lambda_l)\hat{f}(l)\chi_l - \sum_{l=0}^{N-1} g(t\lambda_l)\hat{h}(l)\chi_l \right\| \\
&= \|f - h\|
\end{aligned}$$

□

We can now construct the scattering operation in a cascade fashion in

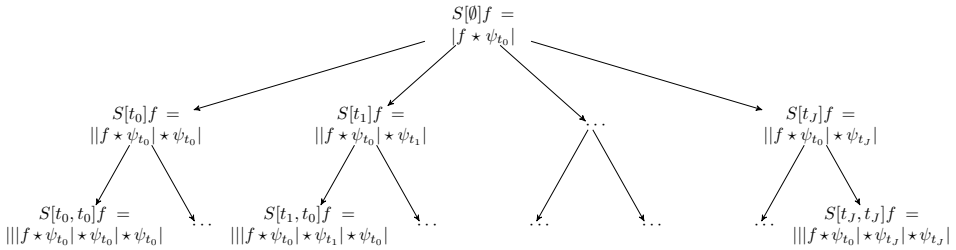


Figure 6.2: Illustration of the convolution network using scattering operation. A function f propagated through the network using SGWT and modulus operation. The scattering coefficients are recursively propagated through the branches by convolution using SGWT in multiple scales, then applied with modulus operation to generate scattering coefficients in the next layer.

multiple layers, which turns out as a convolution network as

$$S[\emptyset]f = |f \star \bar{\psi}_{t_0}| \tag{6.6}$$

$$S[t_{l_1}]f = ||f \star \bar{\psi}_{t_0}| \star \bar{\psi}_{t_{l_1}}| \tag{6.7}$$

$$S[t_{l_1}, t_{l_2}]f = |||f \star \bar{\psi}_{t_0}| \star \bar{\psi}_{t_{l_1}}| \star \bar{\psi}_{t_{l_2}}| \tag{6.8}$$

$$\vdots$$

where $S[l]f$ gives the scattering coefficients obtained by the wavelet bases at the lth layer. In layer l_k, the scattering coefficients from the previous layer l_{k-1} are processed by the forward wavelet transform at multiple scales and then the modulus operator is applied to generate the new scattering coefficients to be passed to the next layer l_{k+1}. This construction of convolution network is illustrated in Fig. 6.2.

Our construction of scattering transform is distinct from the transform proposed in (Bruna and Mallat, 2013) in that we are using scales for branching, not the rotation. In a graph setting, we do not have a notion of direction, but the anisotropic filtering is still achieved by the graph edges.

6.4 Simulation Experiments for Group Analysis

In this section, we demonstrate results of statistical analysis for identifying group level differences using our framework on a population of synthetic images, representative of two distinct groups.

6.4.1 Experimental Design

For the first group, we consider a default 'healthy' image and for the second group, we artificially introduce two separate holes, which is a representative of the disease effect, see top two rows of the first column in Fig. 6.3. Based on these 'models' for the two groups, which serve as the mean μ, we draw samples from each group to populate the cohort: 20 images for each group (total of 40). The image intensities are normalized to $[0,\ 1]$. In other words, for each individual (healthy or diseased) image, the intensity at a specific pixel is drawn from a distribution centered on the mean intensity for *that pixel* in the model image (healthy or diseased). Notice that at this stage, the images are in correspondence and performing a pixel-wise parametric test is meaningful and will reveal precisely the holes as the group-wise difference. Now, we apply a transformation A, which consists of a random rotation and translation drawn from Gaussian distribution (with a mean of 5 percent of the image space for translation, 10° for rotation and a variance 1 for both), to each image to simulate deformations in the images. The pixels across the dataset are *no longer in correspondence*, so a pixel-wise test cannot be performed. We then add Gaussian noise with mean 0 and standard deviation 0.3 to each pixel in each image. The modified images for the healthy group (similarly for diseased) are,

$$Y_{healthy}(i,j) = A\mu_{healthy}(i,j) + N(0, 0.3) \tag{6.9}$$

where A is an arbitrary transformation sampled from a distribution described above. Examples of these synthetic images are shown in the top two rows of the second and the third column of Fig. 6.3.

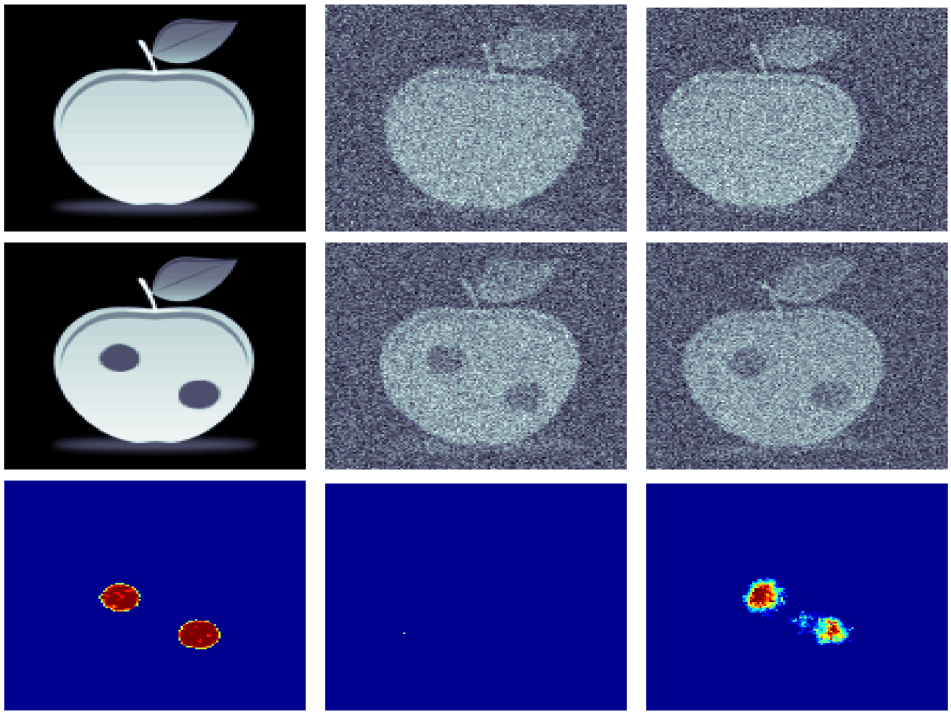


Figure 6.3: Simulation of our framework using synthetic images. First row: the mean image (left) and synthetic images (middle and right) from the normal group, Second row: the mean image (left) and synthetic images (middle and right) from the diseased group , Third row: p-value maps from group analysis using initial images (left), using images with random rotation and translation (middle) and using our framework (right). The bottom row shows that standard statistical analysis detects the group differences correctly on registered images, but it fails in presence of rotation and translation in the images. However, our algorithm accurately detects the group differences (the two holes) despite the deformations.

6.4.2 Results

From the two representative images (top and middle row images in the first column of Fig. 6.3), we can easily tell that the true difference between the models of the groups are the two holes. When standard statistical group analysis is applied pixel by pixel on the two groups of images (with noise and transformation), by performing a t-test at each pixel and using Bonferroni correction at $\alpha = 0.05$ level for multiple comparisons correction, we detect no pixels as showing significant group differences. This is expected because the pixels are no longer in correspondence. On the other hand, using our framework, i.e.,

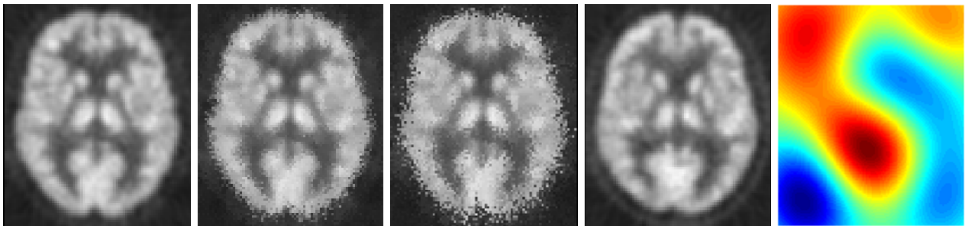
first constructing grid graph using the image intensities, applying scattering transform on the graphs to obtain scattering coefficients and performing a hypothesis test at each pixel, and finally, applying a Bonferroni correction at $\alpha = 0.05$ level, we are able to detect most of the true group differences. This result demonstrates that our framework does offer a reasonable level of invariance to local rotation and translations. If we imagine the two holes in the disease model a consequence of some pathology and the samples drawn from that distribution reflect individual level variations, these simulations suggest that even if the registration is not perfect, the downstream statistical analysis can be made robust to such errors.

6.5 Statistical Analysis on FDG-PET Scans

In this study, we utilize a standard statistical group analysis procedure on real 3-D FDG-PET scans to identify those brain regions that are related to Alzheimer's disease (AD) risk factors, the Apolipoprotein E (APOE) genotype of the subjects and evaluate our framework in the presence of different levels of registration error. Note that an increase in FDG intensity, which is caused by an uptake of glucose in certain brain tissue, is popular for analyzing AD pathology. APOE $\epsilon 4$ is genetically related to the development of AD, where the risk for AD is largely increased with increase in the number of APOE $\epsilon 4$ alleles (Corder et al., 1993).

6.5.1 Dataset

We use partial WRAP dataset of 130 healthy control participants where some subjects have potential AD risk factors (i.e., APOE $\epsilon 4$ genotype, family history and etc). The cohort is comprised of 38 males and 92 females, with a mean age of 64.18. The FDG-PET scans are spatially registered to the Montreal Neurological Institute (MNI) space, and image intensity values are normalized using intensities from the cerebellum as the reference region. During the registration process, the original image is warped to the template image using a deformation field. We introduce errors in this deformation field by 1) adding



(a) (b) (c) (d) (e)

Figure 6.4: Registered FDG-PET scans of a subject. a) Using the original deformation field, b) Using deformation field with 5% noise level, c) Using deformation field with 10% noise level, d) Using deformation field with spatially correlated noise, e) A slice of GRF used for generating d).

uncorrelated errors, up to 10% of the deformation magnitude at each voxel, 2) adding realizations of a Gaussian Random Field (GRF) with variance 0.5 to simulate spatially correlated error (~5 voxels) to simulate imperfect registration. Examples of these images are shown in Fig. 6.4.

6.5.2 Experimental Setup

In this experiment, we used two different datasets with 10% noise level and GRF respectively in the deformation field for registration. As a baseline, we used properly registered FDG-PET images and divided them into two groups using APOE $\epsilon 4$ status, a well-known AD risk factor. We applied *t*-test at each voxel and thresholded the resultant *p*-values at 0.05 level, and the resultant *p*-values were projected onto a T1-weighted template image to identify the risk factor specific regions, which would serve as the ground truth.

On the imperfectly registered datasets, i.e., 1) 5% and 10% noise levels, 2) GRF structured noise in the deformation field, we applied our framework to detect the risk factor specific brain regions. Given an FDG-PET image I for each subject, we first constructed a grid graph (i.e., defined each voxel as a vertex and defined an edge between a voxel and its six neighboring voxels in the 3-D volume space with edge weights of $exp(-\|I(x) - I(y)\|^2/\sigma^2)$ with $\sigma = 0.05$. Then, the proposed scattering transform on graphs was applied on each grid graph to generate scattering coefficients. For the scattering transform, we used a wavelet transform with five scales for each layer, and we went up to the third

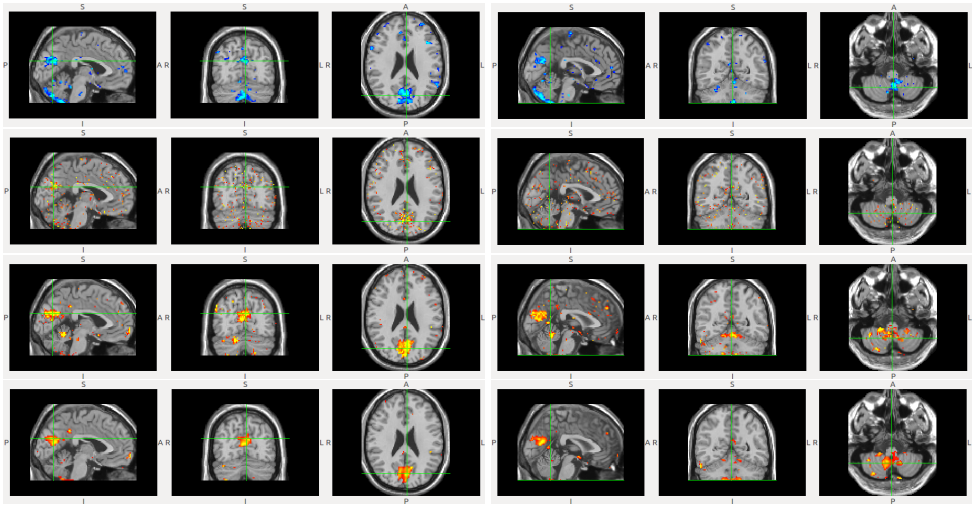


Figure 6.5: Comparisons of statistical group analysis results on cuneus (left column) and cerebellum (right column) using FDG-PET scans with respect to APOE genotype. First row: Using properly registered scans (baseline), Second row: Using imperfectly registered scans (10% noise level), Third row: Our results on imperfectly registered scans (10% noise level), Fourth row: Our results on imperfectly registered scans (GRF noise). Compared to the baseline results in the first row, the results in the second row show decreased statistical power with increased error in the registration, and is unable to detect the cuneus. However, even when there are errors during the registration (voxel-wise noise or GRF), our algorithm correctly identifies the cuneus as shown in the third and fourth row, consistent with the analysis performed on properly registered images.

layer of the cascade to derive scattering coefficients for the original image. In much of brain image analysis, lower frequency components with less noise are preferred, so we took the first three descendant branches from each branch in each layer. Then, these coefficients were defined in a vector form at each voxel (total of 13 features), and Hotelling's T^2 test was applied at each voxel. Again, these resultant p-values were thresholded at 0.05 level and projected onto a T1-weighted template image to obtain a apples-to-apples comparison with the baseline.

6.5.3 Experimental Results on FDG-PET Scans

The result of our analysis is demonstrated in Fig. 6.5. The baseline analysis using APOE $\epsilon 4$ as a predictor on properly registered FDG-PET images revealed

cuneus regions, known to be closely tied to AD pathology, and some lower cerebellum regions as shown in blue-light blue regions in the first row of Fig. 6.5 with p-values in the range of $[0,\ 0.05]$ in $-\log_{10}$ scale. However, as shown in the second row of Fig. 6.5 in red-yellow, the statistical power to detect risk factor related regions diminishes due to noise in the registration process. At the 10% noise level, it returned salt-and-pepper type result, failing to yielding meaningful result. In contrast, using our algorithm on the both datasets with 10% level voxel-wise noise and GRF, we were able to successfully detect exactly the same regions (i.e., cuneus and lower cerebellum regions) that were found in the baseline analysis, as demonstrated in red-yellow regions in the third and fourth row of Fig. 6.5. The results above indicate that even when a good registration is unavailable, this framework may offer a robust solution by performing the same analysis using a deformation invariant representation of the original images.

6.6 Summary

It is well known that the statistical analysis of brain imaging data (including classification and regression experiments) can only proceed once the images have been warped into a common template space. But when individual subjects have significant atrophy or a pathology (resulting from tumor or vascular factors), the registration is imperfect. Separately, in some populations creating a common template itself may be difficult which leads to sub-optimal registration. But independent of where the registration errors come from, their effect on the downstream analysis can be significant and has serious implications on the success of the study. We provide an algorithm that derives local deformation invariant representations of the image. In practice, this means that inference using the image data can proceed as if a (much superior) registration method were available.

Our method is based on the recently proposed Scattering transform which we adapt in interesting ways using ideas related to spectral graph wavelets in the harmonic analysis literature. The performance of the proposed approach does not depend on whether the noise in the warp field is i.i.d. or correlated.

The scattering operator is provably invariant to the action of any compact Lie group on the measured signal (Mallat, 2012). We show that deriving wavelet expansions using the graph representation of the data (together with other modifications) makes the scatter transform a viable tool for analyzing large 3-D image datasets. The proposed ideas have direct applications in neuroimaging analysis but are likely to be more broadly applicable in other computer vision problems where invariance to group actions (e.g., rotation, translation, diffeomorphism) is desired.

7 LATENT VARIABLE GRAPHICAL MODEL SELECTION USING HARMONIC ANALYSIS FOR HUMAN CONNECTOME PROJECT (HCP)

As introduced in Chapter 1, a major goal of brain imaging studies is to identify its associations with covariates such as genotype, risk factors, and so on. For this, a large scale study, Human Connectome Project (HCP), was launched recently to characterize the structural network map of the human brain and find the associations between brain connectivity and various covariates. Unfortunately, the set of image derived measures and the set of covariates in the dataset are both large, so we must first estimate a 'parsimonious' set of relations between the measurements. For instance, a Gaussian graphical model will show conditional independence between the random variables, which can then be used to setup specific downstream analyses. But most such data involve a large list of 'latent' variables that remain unobserved, yet affect the 'observed' variables substantially. Accounting for such latent variables is not directly addressed by standard precision matrix estimation, and is tackled via highly specialized optimization methods. We offer a unique harmonic analysis view of this problem. By casting the estimation of the precision matrix in terms of a composition of low-frequency latent variables and high-frequency sparse terms, we show how the problem can be formulated using a new wavelet-type expansion in non-Euclidean spaces. Our formulation poses the estimation problem in the frequency space and shows how it can be solved by a simple sub-gradient scheme. We provide a set of scientific results on ~500 scans from the recently released HCP data where our algorithm recovers highly interpretable and sparse conditional dependencies between brain connectivity pathways and well-known covariates.

7.1 Overview

Consider a large scale neuroimaging study, e.g., the ongoing Human Connectome Project (HCP), where diffusion weighted magnetic resonance images (diffusion MRI) are acquired for a cohort of participants. Each subject provides a variety of clinical and cognitive measures in addition to the images, as well as demographic information such as age, gender, education status and so on. Such a rich data resource offers an unprecedented opportunity to answer many scientific questions. For instance, how do brain networks differ across gender, and does education or genotype have an association with structural brain connectivity *beyond* the expected effects of age? Until recently, the scientific community had limited means to answer such questions because public datasets were either small, not well curated or the imaging protocols used for acquisition were too heterogeneous. The recent public release of images (and covariates) from the HCP study makes such an analysis possible *if* we can address the associated modeling issues that arise in performing inference on such a high dimensional dataset.

A fundamental scientific goal in statistical analysis of HCP (and similar datasets) is to identify associations between the full set of variables and the entire spectrum of image-derived measurements (Akil et al., 2011; Marcus et al., 2011; Kim et al., 2014, 2015a). For example, are a subset of the clinical covariates highly predictive of the inter-regional connectivity derived from the images? The traditional approach here may proceed by estimating a graphical model that best explains the data: where the nodes correspond to the full set of covariates (image-derived measures *and* clinical/cognitive scores) as jointly Gaussian random variables. By estimating the inverse of the covariance matrix between the variables, we precisely recover the graphical model structure. This may then be used to setup hypothesis driven structural equation models (SEM) (Ullman and Bentler, 2003) or simple regression model based experiments. The difficulty is that in many modern image analysis problems, the total number of such covariates, say p, is far larger than the number of samples (subjects) n in the study. Classical model selection is highly problematic in this high dimensional setting since the empirical statistics are often poorly behaved. The

popular solution here is to impose a *sparsity* regularization on the inverse of the covariance matrix Σ^{-1}. Using a ℓ_1 penalty on the entries of this matrix, under mild conditions, one can guarantee that the maximum likelihood solution will recover the true model (Yuan and Lin, 2007; Friedman et al., 2008). In the last five years, this idea has been extensively used in a broad spectrum of applications in computer vision (Gu et al., 2007; Marlin and Murphy, 2009), machine learning (Banerjee et al., 2008; Raskutti et al., 2008; Liu et al., 2010; Yuan, 2010) and medical imaging (Smith et al., 2011; Varoquaux et al., 2010; Huang et al., 2010).

The formulation above, given its broad applicability, has been heavily studied and we now have a comprehensive treatment of efficient optimization routines (Banerjee et al., 2006; d'Aspremont et al., 2008; Scheinberg et al., 2010; Oztoprak et al., 2012) and regularization properties (Raskutti et al., 2008; Lam and Fan, 2009). These developments notwithstanding, there are various situations in medical image analysis, computational biology and other applications, which are not direct fit for the standard sparse inverse covariance matrix estimation model. For instance, in many real-world studies, there are a non-trivial number of *latent* variables that either *cannot* be directly observed or can only be measured at a high monetary cost or discomfort to the subject. The incorporation of such latent variables in the estimated structural relationship, generally called "latent variable graphical models", is not as extensively studied (Jordan et al., 1999).

Related Work. There is some degree of consensus that a straightforward incorporation of such 'latent' variables in the default construction described above is problematic. Therefore, existing approaches (Dempster et al., 1977) must pre-specify the number of such variables in an ad-hoc manner and proceed with a bi-level non-convex scheme to estimate the parameters. There are other combinatorial heuristics (Elidan et al., 2007) which cluster the observed variables and assign them incrementally to a latent variable. The practical effectiveness of such algorithms varies and they offer few theoretical guarantees. An interesting recent paper (Chandrasekaran et al., 2012) resolves many of these problems and presents an algorithm where all variables (observed and latent) are jointly Gaussian. The main idea is to approximate the sample

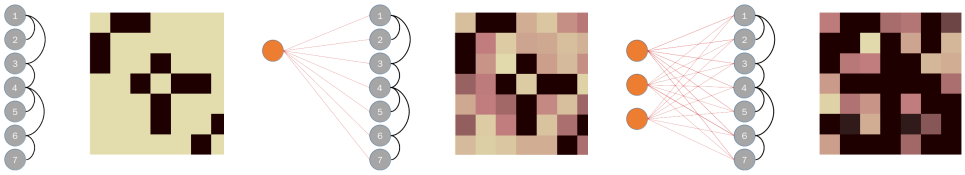


Figure 7.1: Demonstration of latent variables (orange nodes) affecting the observation (7 grey nodes) leading to dense precision matrix. Left: no latent variable and corresponding true relationship and sparse precision matrix, Middle: one latent variable and corresponding dense precision matrix, Right: many latent variables and corresponding denser precision matrix

covariance matrix Σ in a way where the corresponding Σ^{-1} is expressed as a sum of a sparse matrix and a low-rank matrix. This recovers the influence of the unobserved latent variables as well as the conditional graphical model, as desired. This strategy works well as long as the low rank requirement remains valid; however, as the number of latent variables grow, the data may deviate farther from the low rank assumption. Consequently, the sparse term must explain a larger 'mass' of the data and the estimated matrix becomes denser.

The above discussion suggests that the means of regularizing the degrees of freedom (i.e., the low rank term) for the *latent components* may be not be ideal from a numerical perspective as the number of latent variables grow. Notice that the literature suggests that *high rank matrix completion* (columns of the matrix belong to a union of multiple low-rank subspaces) uses a set of concepts that are quite different from those used for completing low-rank matrices (e.g., nuclear norm). So, a potential solution in our graphical model setting must also look for alternatives to the algebraic characterization (used in (Chandrasekaran et al., 2012)). Certain classical tools at the high level, express a closely related intuition. Consider the following simple idea. If we think of the precision matrix as the composition of *low* and *high* frequency terms, the lower order terms may easily serve as a proxy for the latent components. Then, by asking that the remaining contribution should be sparse yields a similar overall effect as (Chandrasekaran et al., 2012) but does not directly involve spectral relaxations of the rank constraint. Fortunately, harmonic analysis offers a natural tool for such needs via wavelets which are used in the previous chapters.

Adopting the ideas above, our proposal here is to perform multi-resolution analysis of the main component for graphical model selection (i.e.,sample precision matrix), using the formulation of mother wavelets in non-Euclidean spaces introduced in Chapter 2. Analyzing different bandwidths (i.e.,scale) in a dual space (i.e.,frequency space) derived from a covariance matrix, we may be able to find the optimal band that yields a desirable property for the estimated precision matrix. Later in the experimental section, we demonstrate that such a framework provides a good solution to a sparse precision matrix estimation problem. When there exist (potentially a large number of) unknown latent variables, our method still works well while baseline methods fail to achieve the "sparsity" in the estimated precision matrix.

7.2 Harmonic Analysis of Latent Variable Graphical Models

Using the wavelet concepts in non-Euclidean spaces described in Section 2.2 in hand, we will describe our formulation for estimating a precision matrix while concurrently taking into account the effect of an unknown but large number of latent components. Our procedure below will parameterize the to-be-inferred graphical model *not* in terms of its precision matrix directly, rather via its low and high frequency components. Operating on these latent (low-frequency) and sparse (high-frequency) pieces will model the structural associations within the graph. Recall that recent developments in wavelet analysis on discrete spaces such as graphs have overwhelmingly been used to analyze signals defined on the nodes where the graph has a "fixed" (known) structure. In order to apply wavelet analysis to our problem, we will need to introduce a few key technical results that are summarized below, and described in detail in this section. **(a)** First, we will introduce multi-resolution analysis for modeling the graph *structure* and *not* just the measurement at individual graph nodes. We will define a new set of basis functions for estimating the graph structures and provide theoretical conditions which guarantee its validity. **(b)** Second, we will introduce an information theoretic "closeness" measure for

graph structure (i.e., precision matrices). Here, we will identify an additional condition which will yield a valid symmetric positive definite matrix at each scale s. **(c)** Finally, we will discuss our optimization scheme in the dual space (i.e., frequency domain) with a simple gradient descent method.

7.2.1 Multiscale Analysis of a Precision Matrix

Let us assume we are given a positive definite covariance matrix Σ of size $n \times n$. Now, Σ can be easily decomposed in terms of its eigenvector and eigenvalues as,

$$\Sigma = V \Lambda V^T = \sum_{\ell=1}^{n} \lambda_\ell V_\ell V_\ell^T \tag{7.1}$$

where the ℓth column vector of V is the ℓth eigenvector and the ℓth diagonal of Λ is the corresponding ℓth eigenvalue of Σ which are all positive. Then, the precision matrix Θ is given as the inverse of the covariance matrix as

$$\Theta = \sum_{\ell=1}^{n} \frac{1}{\lambda_\ell} V_\ell V_\ell^T = \sum_{\ell=1}^{n} \sigma_\ell V_\ell V_\ell^T \tag{7.2}$$

where $\sigma = \frac{1}{\lambda}$ and σ are positive since λ are positive. Notice that both Σ and Θ are positive definite and self-adjoint, so their eigenvectors can be used for defining a Fourier type of transform which is analogous to the graph Fourier transform as in (2.8). For multi-resolution analysis of the precision matrix Θ, we first define our basis functions as

$$\psi_{\ell,s}(i,j) = g(s\sigma_\ell) V_\ell^*(i) V_\ell(j), \forall \ell \in \{1, \ldots, n\} \tag{7.3}$$

at scale s and along the ℓth basis. Since we deal only with real valued functions, to avoid notational clutter, we will omit the conjugate operation for the eigenfunctions, i.e., $V^*(i) = V(i)$. These basis functions are analogous to mother wavelets and yield a nice result which we will present shortly. Now, we can easily setup a transform of the precision matrix using our basis above. This

yields wavelet-like coefficients as

$$\begin{aligned} W_{\Theta,s}(\ell) &= \langle \Theta, \psi_{\ell,s} \rangle \\ &= \sum_{i,j}^{n,n} \sum_{\ell'=1}^{n} \sigma_{\ell'} V_{\ell'}(i) V_{\ell'}(j) g(s\sigma_\ell) V_\ell(i) V_\ell(j) \\ &= \sigma_\ell g(s\sigma_\ell). \end{aligned} \tag{7.4}$$

Using $W_{\Theta,s}(\ell)$, the multi-resolution reconstruction with a non-constant weight ds/s is obtained by

$$\tilde{\Theta}(i,j) = \frac{1}{C_g} \int_0^\infty \frac{1}{s} \sum_{\ell=1}^{n} W_{\Theta,s}(\ell) \psi_{\ell,s}(i,j) ds. \tag{7.5}$$

Roughly speaking, this can be viewed as the weighted average of multi-resolution reconstruction over scale s.

A natural question here is whether we can guarantee if the reconstruction in (7.5) is identical to the original precision matrix Θ. To address this issue, we define the *admissibility condition* for the function defined on the structure (or *edges*) of the graph. A kernel $g(x)$ is said to satisfy the *admissibility condition* if the following condition holds

$$\int_0^\infty \frac{g^2(x)}{x} dx =: C_g < \infty \tag{7.6}$$

when the reconstruction is defined with a non-constant weight dx/x as (7.5). Lemma 7.1 below shows that using the bases we constructed in (7.3), if $g(x)$ satisfies *admissibility condition*, the matrix reconstruction in (7.5) is *identical*, namely, $\tilde{\Theta}(i,j) = \Theta(i,j)$.

Lemma 7.1. *If* $\Theta \succ 0, \Theta = \Theta^T$ *and kernel* g *satisfies the admissibility condition*

$$\int_0^\infty \frac{g^2(s\sigma)}{s} ds =: C_g < \infty \tag{7.7}$$

then,

$$\frac{1}{C_g}\int_0^\infty \frac{1}{s}\sum_{\ell=1}^{n} W_{\Theta,s}(\ell)\psi_{\ell,s}(i,j)ds = \Theta(i,j) \tag{7.8}$$

Proof. By substituting $\psi_{\ell,s}$ and $W_{\Theta,s}(\ell)$ with (7.3) and (7.4) respectively in graph Fourier basis, the left hand side of (7.8) is given as

$$\begin{aligned}
&\frac{1}{C_g}\int_0^\infty \frac{1}{s}\sum_{\ell=1}^{n} W_{\Theta,s}(\ell)\psi_{\ell,s}(i,j)ds \\
&= \frac{1}{C_g}\int_0^\infty \frac{1}{s}\left(\sum_{\ell=1}^{n} \sigma_\ell g(s\sigma_\ell)g(s\sigma_l)V_\ell(i)V_\ell(j)\right) ds \\
&= \frac{1}{C_g}\int_0^\infty \frac{1}{s}\left(\sum_{\ell=1}^{n} \sigma_\ell g^2(s\sigma_\ell)V_\ell(i)V_\ell(j)\right) ds \\
&= \sum_{\ell=1}^{n}\left(\frac{1}{C_g}\int_0^\infty \frac{g^2(s\sigma_\ell)}{s}ds\right)\sigma_\ell V_\ell(i)V_\ell(j) \\
&= \sum_{\ell=1}^{n} \sigma_\ell V_\ell(i)V_\ell(j) = \Theta(i,j) \text{ by the assumption in (7.7)}
\end{aligned}$$

□

We can derive a stronger result showing that using the bases in (7.3), the admissibility condition holds for two parameter kernels as well, i.e., $g(s,\sigma)$. This allows defining a kernel, if desired, that separately handles the influence of the eigenvalue σ and a scale parameter s.

Lemma 7.2. *If kernel* g *satisfies the admissibility condition*

$$\int_0^\infty \frac{g^2(s,\sigma)}{s}ds =: C_g < \infty \tag{7.9}$$

then,

$$\frac{1}{C_g}\int_0^\infty \frac{1}{s}\sum_{\ell=1}^{n} W_{\Theta,s}(\ell)\psi_{\ell,s}(i,j)ds = \Theta(i,j) \tag{7.10}$$

Proof. Note that since $\int_0^\infty \frac{g^2(s\sigma)}{s} ds = \int_0^\infty \frac{g^2(x)}{x} dx =: C_g$, the admissibility condition can be written in both ways.
Using the definition of $\psi_{\ell,s}$ and $W_{\Theta,s}(\ell)$ for the reconstruction of Θ,

$$\begin{aligned}
&\frac{1}{C_g}\int_0^\infty \frac{1}{s}\sum_{\ell=1}^{n} W_{\Theta,s}(\ell)\psi_{\ell,s}(i,j)ds \\
&\qquad = \frac{1}{C_g}\int_0^\infty \frac{1}{s}\left(\sum_{\ell=1}^{n} \sigma_\ell g(s\sigma_\ell) g(s\sigma_l) V_\ell(i) V_\ell(j)\right) ds \\
&\qquad = \frac{1}{C_g}\int_0^\infty \frac{1}{s}\left(\sum_{\ell=1}^{n} \sigma_\ell g^2(s\sigma_\ell) V_\ell(i) V_\ell(j)\right) ds \\
&\qquad = \sum_{\ell=1}^{n}\left(\frac{1}{C_g}\int_0^\infty \frac{g^2(s\sigma_\ell)}{s} ds\right)\sigma_\ell V_\ell(i) V_\ell(j) \\
&\qquad = \sum_{\ell=1}^{n} \sigma_\ell V_\ell(i) V_\ell(j) = \Theta(i,j) \text{ by the admissibility condition in (7.8)}
\end{aligned}$$

□

This two parameter kernel result can be used for functions defined on either nodes (commonly used in non-Euclidean Wavelets) or edges (graph structure). The admissibility condition for the classical SGWT (one parameter kernel for the functions defined on nodes) is studied in (Hammond et al., 2011) and consistent with our result. Based on this harmonic analysis of graphical models, we next describe our main estimation algorithm to recover the sparse precision matrix by explicitly taking into account the contribution of the latent components.

7.3 Estimating the Optimal Scale for a Sparse Precision Matrix

In this subsection, we describe the optimization scheme to estimate $\tilde{\Theta}$ which satisfies two properties: **i)** it is consistent with the empirical Θ and **ii)** satisfies sparsity properties (in the sense of the multi-resolution characterization)

The reconstruction of Θ at level s is given by

$$\tilde{\Theta} = \sum_{\ell=1}^{n} \sigma_\ell g^2(s\sigma_\ell) V_\ell V_\ell^T = \sum_{\ell=1}^{n} K(s, \sigma_\ell) V_\ell V_\ell^T \tag{7.11}$$

where $K(s, \sigma_\ell) := \sigma_\ell g^2(s\sigma_\ell)$. To keep notations concise, we will often use K as shorthand in this subsection. To perform the reconstruction at every level s, the kernel function g should satisfy the condition, $g^2(x) > 0, \forall x \geqslant 0$. Then, one can easily check that $\tilde{\Theta}$ is symmetric positive definite, i.e., $\tilde{\Theta} \in \text{SPD}$, exactly as desired.

At a high level, we seek for a $\tilde{\Theta}$ which is similar to the empirical (potentially non-sparse) estimate, Θ. To do so, we need to define "closeness" between our estimate $\tilde{\Theta}$ and Θ. We consider the two matrices as corresponding Gaussian distributions with zero mean, but with covariance matrices $\tilde{\Sigma}$ and Σ.

Using KL-divergence $\text{KL}(\cdot\|\cdot)$ between the two Gaussian densities, we can measure "closeness" by

$$\text{KL}(p(x; \tilde{\Sigma}) \| p(x; \Sigma)) = \frac{1}{2} D_{ld}(\Sigma, \tilde{\Sigma}) = \frac{1}{2} D_{ld}(\tilde{\Theta}, \Theta) \tag{7.12}$$

The last two identities express closeness by Bregman divergence using the log determinant (Davis et al., 2007). Given two matrices A and A_0, the Bregman divergence between the two matrices $D_{ld}(A, A_0)$ is given as

$$D_{ld}(A, A_0) = \text{tr}(AA_0^{-1}) - \text{logdet}(AA_0^{-1}) - n. \tag{7.13}$$

With this fidelity measure, our objective is to find the optimal scale s which minimizes the Bregman divergence using $\text{logdet}(\cdot))$ between the empirical precision matrix Θ and the sparse reconstruction $\tilde{\Theta}$. We impose a sparsity penalty in the usual way using the ℓ_1-norm of the matrix. Then, our optimization

problem is given as,

$$\min_{s \geqslant 0} \ \operatorname{tr}(\tilde{\Theta}\Theta^{-1}) - \operatorname{logdet}(\tilde{\Theta}\Theta^{-1}) - n + \gamma|\tilde{\Theta}|_1 \tag{7.14}$$
$$\text{subject to } \tilde{\Theta} = \sum_{\ell=1}^{n} \sigma_\ell g^2(s\sigma_\ell) V_\ell V_\ell^T.$$

Substituting in the identity from (7.1) for Θ, we obtain an almost unconstrained optimization model (which only involves one non-negativity constraint),

$$\begin{aligned} \min_{s \geqslant 0} \sum_{\ell=1}^{n} \lambda_\ell K(s, \sigma_\ell) - \sum_{\ell=1}^{n} \log(\lambda_\ell K(s, \sigma_\ell)) - n \\ + \gamma \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \sum_{\ell=1}^{n} K(s, \sigma_\ell) X_\ell(i, j) \right| \end{aligned} \tag{7.15}$$

where $X_\ell = V_\ell V_\ell^T$ and $X_\ell(i, j)$ is i, jth element in X. The optimal (sparse) precision matrix will then correspond to some s which minimizes (7.14) or (7.15).

7.3.1 Deriving the First Derivative for Optimization

To optimize (7.14), we compute the first derivative of D with respect to s, which can be written as

$$\frac{d}{ds}\operatorname{tr}(\tilde{\Theta}\Theta^{-1}) - \frac{d}{ds}\operatorname{logdet}(\tilde{\Theta}\Theta^{-1}) + \frac{d}{ds}\gamma|\tilde{\Theta}|_1 \tag{7.16}$$

Here, we calculate $\frac{d}{ds}\operatorname{tr}(\sum_{\ell=1}^{n} \lambda_\ell K(s, \sigma_\ell) V_\ell V_\ell^T)$ taking a derivative of each element and then taking the sum of the diagonal elements of $\tilde{\Theta}\Theta^{-1}$ and we obtain,

$$\frac{d}{ds}\operatorname{tr}(\sum_{i=1}^{n} \lambda_\ell K(s, \sigma_\ell) V_\ell V_\ell^T) = \sum_{\ell=1}^{n} \lambda_\ell K'(s, \sigma_\ell) \tag{7.17}$$

where $K'(s, \sigma_\ell) := \partial K / \partial s$. The derivative of the second term takes the form,

$$\frac{d}{ds}\operatorname{logdet}(\tilde{\Theta}\Sigma) = \sum_{i=1}^{n} \frac{K'(s, \sigma_i)}{K(s, \sigma_i)} \tag{7.18}$$

Notice that the third term involves the ℓ_1 norm which is not differentiable, so we approximate its search direction instead as

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \text{sign}(\tilde{\Theta}(i,j)) \sum_{\ell=1}^{n} K'(s,\sigma_\ell) X_\ell(i,j). \tag{7.19}$$

Combining all three terms together yields a direction to optimize (7.14). The actual optimization then only involves a simple gradient descent-like method.

Remarks. Observe that a precision matrix always has non-zero diagonal elements. So, the sparsity regularization may not be meaningful for diagonal elements. One can impose sparsity for only the off-diagonal elements with minor changes in the third term (7.15) and its search direction (7.19), namely, $\sum_{i\neq j} |\tilde{\Theta}_{(i,j)}|$ and its search direction is

$$\sum_{i\neq j}^{n,n} \text{sign}(\tilde{\Theta}(i,j)) \sum_{\ell=1}^{n} K'(s,\sigma_\ell) X_\ell(i,j). \tag{7.20}$$

7.3.2 Choice of the Kernel Function

For the choice of kernel function $g()$, we used the popular Gaussian function $\exp(-\frac{1}{2}sx)$. This kernel function models diffusion or a random walk process (Spitzer, 2013), and is used to define diffusion type of wavelets (Coifman and Maggioni, 2006; Hou and Qin, 2013). The kernel function itself may not satisfy the admissibility condition, which is not an issue because we work with a single scale estimation. However, to work with all scales concurrently, we will be limited to only that class of kernels which directly satisfies the admissibility condition.

$$K(s,\sigma) = \sigma g^2(s\sigma) = \sigma e^{-s\sigma} \tag{7.21}$$

is able to perfectly reconstruct the original sample precision matrix Θ. Intuitively, in our optimization problem, we should be able to reconstruct the exact Θ in the non-regularized setting. This is because of σ in front of $g()$ in (7.21).

That is, our estimation $\tilde{\Theta}$

$$\tilde{\Theta} = \sum_{l} \sigma_l g^2(s\sigma_l) V_l V_l' \tag{7.22}$$

becomes P when $s = 0$ as

$$\tilde{\Theta} = \sum_{l} \sigma_l e^{-s\sigma} V_l V_l' \tag{7.23}$$

$$= \sum_{l} \sigma_l V_l V_l' = \Theta. \tag{7.24}$$

7.4 Experimental Results for Latent Variable Graphical Model Selection

We demonstrate two sets of experiments, one on synthetic brain network data to validate our framework where the ground truth is available, and the other on the Human Connectome Project (HCP) data. The first experiment evaluates precision matrix estimation results using our framework by comparing it to the estimations from other methods and the ground truth. In the second experiment, we analyze an exquisite recently released imaging dataset of ~ 500 (and increasing) individuals from the Human Connectome Project. We obtain brain connectivity pathways by processing Diffusion Tensor Images (DTI) and analyze this connectivity data jointly with a rich set of covariates. Among the many inter-regional fiber bundles, we focus our analysis on 17 major connections and identify which of the covariates are statistically associated with these major pathways. In both experiments, the objective is to estimate true dependencies between the observed variables when the latent variables are unobserved.

7.4.1 Statistical Dependency Estimation on Synthetic Brain Connectivity Data

In this section, we demonstrate results of precision matrix estimation using synthetic brain connectivity data. Consider a case where we observe a set

of $n_p + n_c$ random variables, i.e., a set of n_p structural brain connections (i.e., pathways) and n_c covariates. We add additional n_h number of latent variables that are assumed to be *unobserved* but statistically influence the full set of observed variables. Then, the statistical model estimation task is to find the true conditional dependencies between the observed variables alone, i.e., properly taking into account the effect of latent factors. In other words, we want to identify which brain connections are statistically associated with the covariates as well as how these pathways are related to one another.

Experimental Design

We set $n_p = 50$ and $n_c = 10$, so the total number of observed variables $n_o = 60$. We run multiple replicates, each for a different setting for the number of latent variables n_h. The dependencies between the brain connections are arbitrarily chosen such that 5% of the elements in the true precision matrix (i.e., ground truth) are non-zeros. We set each covariate to be dependent on the brain connections in a pattern (i.e., first five connections depend on the first covariate, next five connections depend on the second covariate, etc.). These dependencies between observed variables are the ground truth and can be visual checked to see if we estimate the same pattern. The latent variables are then connected to all observed variables with random weights; this ensures that our measurements of the observed variables include an effect from all latent variables. This yields a precision matrix Θ of size $n \times n$ and its corresponding covariance matrix Σ. Synthetic data are sampled from a multi-variate normal distribution using Σ. We draw samples only from the observed variables to construct a sample covariance matrix Σ_o, which serves as the input for estimating $\tilde{\Theta}_o$.

Results

Experimental results with $n_h = 5$ and $n_h = 10$ latent variables are shown in the top/bottom blocks of Fig. 7.2 respectively. Fig. 7.2 (top row) shows in each block the estimated dependencies between the full set of connections (covariates are not shown in the top row). The small spheres represent the

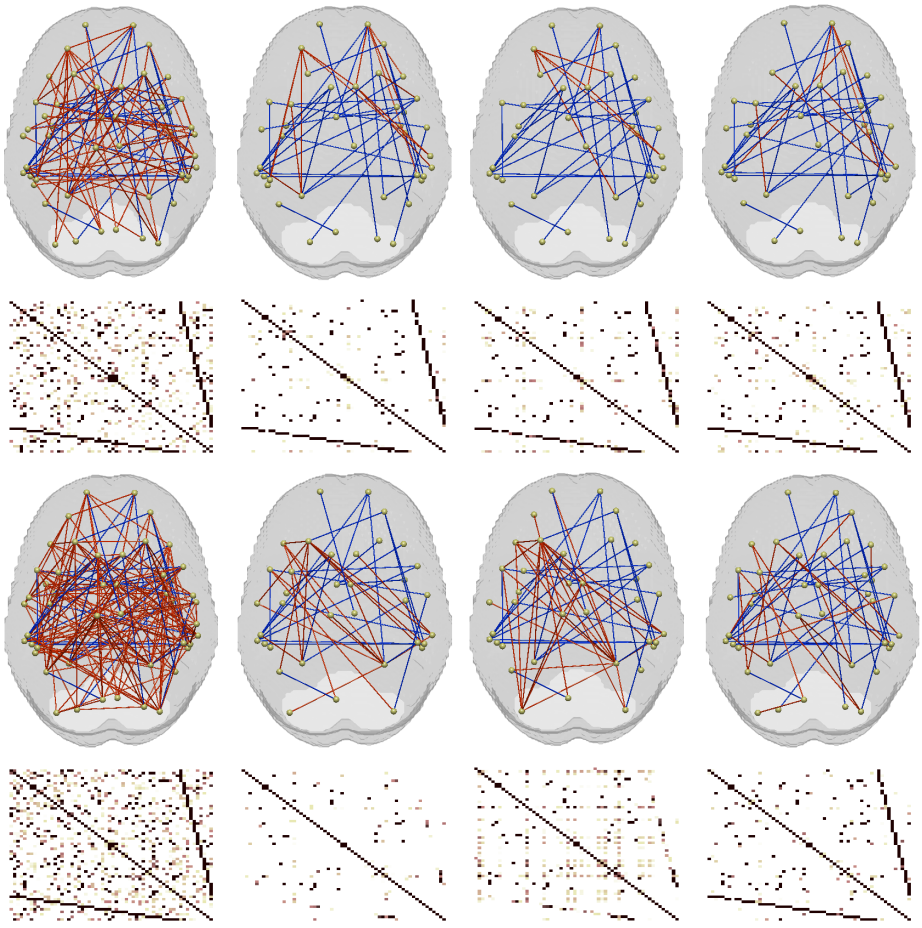


Figure 7.2: Comparison of results from estimation of statistical dependencies between observed variables (when there are at least a few latent components) using synthetic brain network data. Top/Bottom blocks show results for 5 and 10 latent variables respectively, and the top/bottom rows show estimated dependencies in the data (correct estimation in blue and false positive in red) and corresponding precision matrices. First column: sample precision matrix, Second column: result using GLasso, Third column: result using Chandrasekaran et al. (2012), Fourth column: our result. We can observe that while the sample precision matrix is dense, the results in the second, third and fourth column show sparse and more accurate results.

physical centers of each brain connection and the edges in blue/red denote correct/incorrect estimation of the conditional dependency. The bottom row shows the estimated $\tilde{\Theta}_o$ including the covariates. In each block, the first column shows Σ_o^{-1} as the estimated precision matrix $\tilde{\Theta}_o$ (i.e., sample precision matrix).

Here, both sample precision matrices are *dense* due to the effect from the latent variables, leading to a solution which is far from Θ_o. In the second and third column in both blocks, we include results from graphical Lasso Friedman et al. (2008) and the method from Chandrasekaran et al. (2012). When $n_h = 5$, the sparsity pattern in the estimated precision matrices for both baselines and our algorithm are quite similar to the ground truth (few red edges). In the second row, we see that the oblique patterns expressing the relationship between the connectivity and the covariates is also recovered. When the number of latent variables grows, the low rank assumption in Chandrasekaran et al. (2012) becomes weaker and the data deviates from the assumptions of graphical lasso (which assumes all variables are observed). For $n_h = 10$, neither of the baselines are able to recover the conditional dependencies between the connections and covariates (oblique patterns in the precision matrix). On the other hand, the fourth columns (second row) shows that our algorithm recovers Θ_o with a sparsity pattern that is highly consistent with the ground truth Θ_o.

In the following, we demonstrate additional analysis of precision matrix estimation with varying number of latent variables. In Fig. 7.3, estimation results from experiments with different number of latent variables (i.e., 0, 5, 10, 20) are displayed, where the top row in Fig. 7.3 shows the inverse covariance matrix corresponding to different number of latent variables and the bottom row shows our estimations. At a glance, we can easily see that the increase in the number of latent variables makes the inverse covariance matrix denser, on the other hand, our estimation results yield correct sparse graphical models despite the increase. In Fig. 7.4, the precision matrix estimation in different scales with 5 latent variables are demonstrated whose a), b) and c) are the estimation results with $s = 0, 0.5, 0.7$, and d) shows a result with the optimal scale $s = 0.2089$ obtained using our framework. The estimation result significantly varies depending on the scale parameter with dense non-zero elements, but our method is able to find the optimal scale that provides an estimation that is sparse and similar to the sample precision matrix.

Limitations. These results suggest that our algorithm is effective in identifying the true precision matrix even when there are diffusive effects of un-

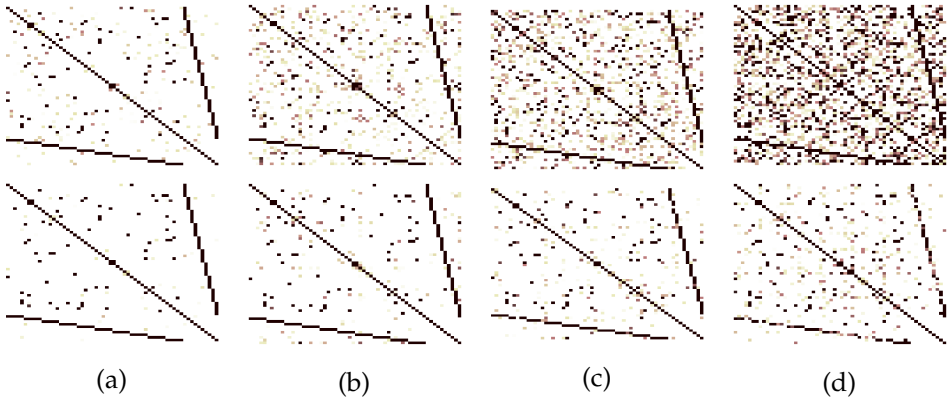


Figure 7.3: Precision matrix estimation with different numbers of latent variables. Top row shows the inverse covariance matrix, and the bottom row shows our estimation result. a) no latent variables, b) 5 latent variables, c) 10 latent variables, d) 20 latent variables. We can easily see that the inverse covariance matrix becomes denser as the number of latent variables increases, while our method yields good estimation of the sparse graphical model.

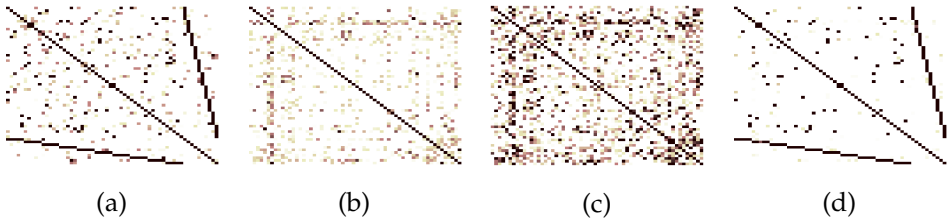


Figure 7.4: Precision matrix with 5 latent variables in different scales. a) estimation with $s = 0$, b) estimation with $s = 0.5$, c) estimation with $s = 0.7$, d) estimation with optimal scale $s = 0.2089$. Our optimization scheme find the optimal scale that gives a sparse graphical model.

observed variables. However, in situations where we have a large number of latent variables and each affects only a small number of observed variables (i.e., high-frequency effect), our algorithm may not be able to identify the correct associations.

7.4.2 Experiments on Human Connectome Project (HCP) Data

In this section, we demonstrate experimental results on a real brain imaging dataset from HCP. The dataset has many covariates that may correspond to latent variables, and we try to estimate relationships among different neural

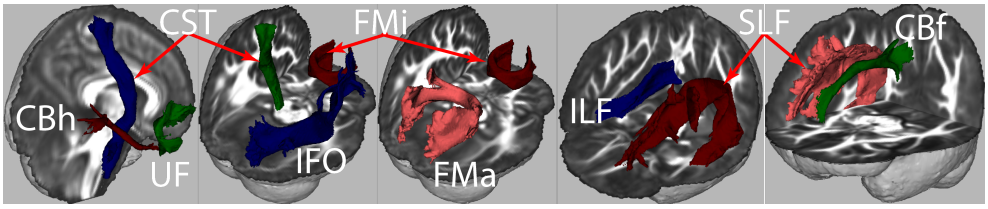


Figure 7.5: The subset of white matter tract fibers used in our experiment (from the seventeen presented in Table 7.1) that are statistically associated to non-imaging covariates.

fiber pathways and important covariates related to demographics, physical health, memory, cognitive ability. By applying our framework on the data, we expect to identify the true relationships by estimating a sparse precision matrix.

HCP Dataset

The HCP[1] project recently made available high-quality imaging and clinical data ((Uğurbil et al., 2013; Glasser et al., 2013)) for over 500 healthy adults (Hodge et al., 2015). We analyzed the high angular resolution diffusion MR imaging (dMRI) dataset, consisting of 489 images (Sotiropoulos et al., 2013; Van Essen et al., 2013).

Imaging Data. We obtained DTI from the dMRI data via standard fitting procedures which were then spatially normalized (Zhang et al., 2006a). Seventeen major white matter connectivity measures were obtained by registering (using ANTS) the publicly available IIT atlas Varentsova et al. (2014) to the HCP template. The average fractional anisotropy (FA) in each pathway was a proxy for the connection strength. Table 7.1 lists the seventeen connectivity pathways.

Non-imaging covariates. Besides the imaging data, HCP provides several categories of non-imaging covariates for the subjects Herrick et al. (2014) covering factors such as cognitive function, demographic variables, education

[1]Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Connection label	Description (count)
Forceps major (FMa)	inter-hemispheric (1)
Forceps minor (FMi)	inter-heispheric (1)
Fornix	inter-hemispheric (1)
Cingulum bundle frontal (CBf)	bi-lateral (2)
Cingulum bundle hippocampal (CBh)	bi-lateral (2)
Cortico-spinal tracts (CST)	bi-lateral (2)
Inferior fronto-occipital (IFO)	bi-lateral (2)
Inferior longitudinal fasciculus (ILF)	bi-lateral (2)
Superior longitudinal fasciculus (SLF)	bi-lateral (2)
Uncinate fasciculus (UF)	bi-lateral (2)

Table 7.1: Pathways spanning connections between all major lobes of the brain (frontal, parietal, occipital and temporal) with several important regions such as amygdala, hippocampus, pre-frontal cortex.

and so on. In our experiments, we chose 22 variables related to demographics, physical health, sleep, memory, cognitive flexibility and other as listed in Table 7.2. These covariates span a wide range high-level human behavior and highly relevant physiological measurements.

Sparse Precision Matrix Estimation on HCP Dataset

Figures 7.5 summarize the results of our experiments on the HCP data. The matrix shown in Fig. 7.6 lists the full set of connections and covariates used in our analysis, along the axes. Our goal was to recover a sparse (and interpretable) precision matrix explaining the conditional dependencies among these variables. It is clear from the figure that our algorithm indeed finds a parsimonious set of statistical relations, among the non-imaging covariates, among the brain pathways as well as across these two groups of variables. As we can expect, several connectivity pathways seem to be involved in several different categories of behavioral measures. Note that similar to the simulation setup, in this case, results from the baseline algorithms were non-sparse and hence harder to interpret. Part of the reason is that none of the measurements were controlled for various (observed or unobserved) nuisance variables. One advantage of our algorithm is to take into account the effect of such latent

Category	**C**ovariates
Demographics	Age, gender, years of education completed (Edu)
Physical health	Height, weight
Alertness	Mini mental status exam (MMSE)
Sleep	Pittsburgh sleep questionnaire (PSQI)
Episodic memory	Picture sequence recall (PicSeq)
Cognitive flexibility	Picture matching accuracy and reaction time (CardSort)
Inhibition	Flanking accuracy and reaction time (Flanker)
Fluid intelligence	Correct responses in Penn progressive matrices (PMA_CR)
Reading	NIH toolbox reading recognition test (ReadEng)
Vocabulary	NIH toolbox picture vocabulary (PicVocab)
Processing speed	NIH toolbox pattern comparison speed (ProcSpeed)
Spatial orientation	Expected number of correct clicks (VSPLOT_CRTE), total off positions (VSPLOT_OFF)
Sustained attention	Short Penn continuous performance test: sensitivity (SCPT_SEN), specificity (SCPT_SPEC), longest run of non-responses (SCPT_LRNR)
Episodic memory	Penn word memory test: total correct responses (IWRD_TOT), reaction time (IWRD_RTC)
Working memory	NIH toolbox sorting working memory (ListSort)

Table 7.2: Full list of non-imaging covariates used in our analysis spanning a wide range high-level human behavior and highly relevant physiological measurements.

nuisance variables automatically.

Finally, since there is no 'ground truth' available for these results, we checked if our findings are corroborated by independent results in the literature. We found that many of the associations in Figs. 7.6 appear as standalone findings in multiple papers (Penke et al., 2012; Booth et al., 2013). For example, the association between the cingulum bundle and processing speed was the focus of (Nestor et al., 2007), whereas (Karlsgodt et al., 2008; Kubicki et al., 2005) identified a relation between longitudinal fasciculus and cognitive/verbal ability and (Zarei et al., 2011) demonstrated that forceps major and gender were related. Significant associations have also been found between integrity of the uncinate fasciculus and spatial working memory (Davis et al., 2009). This is not definitive evidence that we identify the real underlying precision matrix, but promising that most of the identified associations have precedence in the literature.

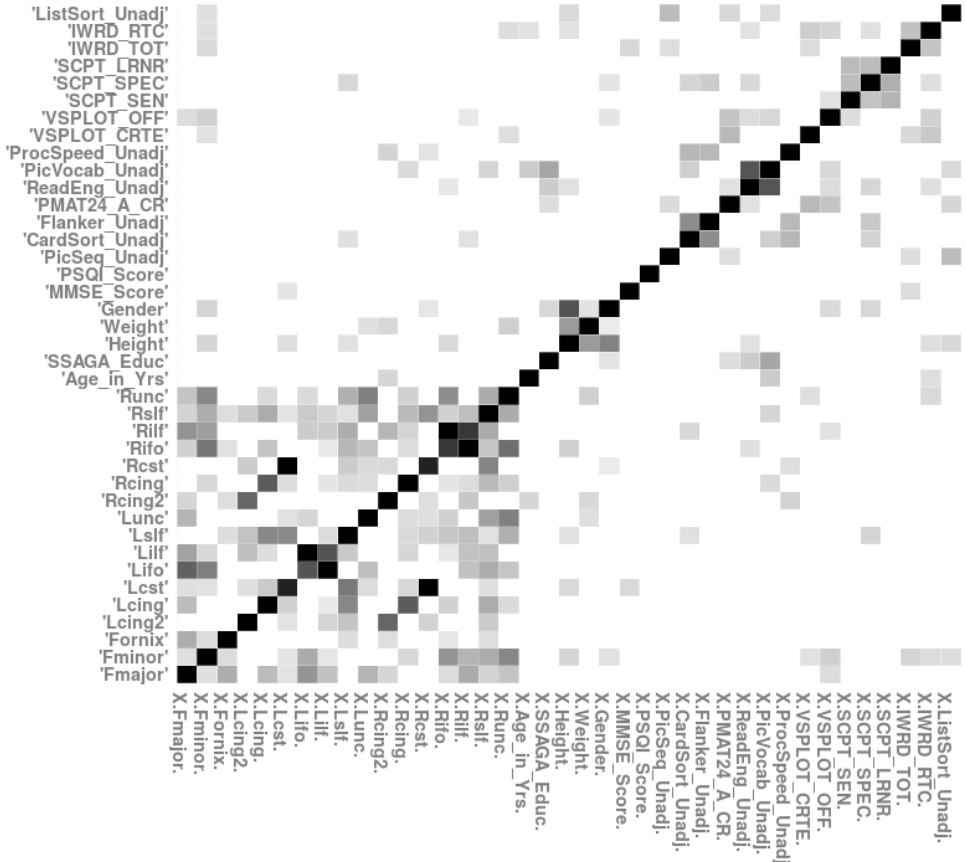


Figure 7.6: Estimated sparse precision matrix on HCP dataset. Notice that the matrix shows sparse connections between the pathways and covariates.

7.5 Summary

Undirected graphical models are used to address a variety of needs in computer vision and machine learning. While existing methods for estimating statistical conditional independence between a set of random variables are quite effective, this analysis becomes problematic when there are multiple latent (unobserved) variables that non-trivially affect our measurements of the observed variables. This situation is becoming more frequent in many modern medical image analysis and computer vision datasets, where the latent variables cannot be measured due to cost or privacy reasons. We propose a

novel perspective on this sparse inverse covariance matrix estimation problem involving latent variables using non-Euclidean wavelet analysis. The experimental results using synthetic brain network data demonstrated that our algorithm provides substantial improvement over other graphical model selection methods. Also, we presented an extensive set of results on the recently released HCP imaging data set showing statistical dependencies between brain connectivity pathways and cognitive/behavioral covariates, and the results that we found are consistent with independent findings in the neuroscience literature.

8 DISCUSSION AND FUTURE WORK

8.1 Summary of Contributions

The work introduced in this thesis tackles fundamental problems that routinely arise in typical neuroimaging studies. The problems described in each chapter were motivated in the context of neuroimaging analysis for Alzheimer's disease and solved using Harmonic Analysis on Graphs. The proposed methods and applications mostly fall into one of the four scopes: 1) developing a sensitive method for analyzing neuroimaging data on graphs to detect subtle group differences, 2) developing a method to compute differences between images (e.g., longitudinal images) even in the presence of systematic variation, 3) developing an image representation that is invariant to local deformation error in brain image registration, 4) developing a graphical model selection method with unknown latent variables to identify relationships between brain regions and covariates. Although these frameworks were adopted to solve interesting problems and evaluated in AD analysis, they can be applied to other domains as long as the data is represented with a structure, i.e., a graph. In the following, we briefly summarize the key contributions of the proposed methods in this thesis.

8.1.1 Sensitive Methods for Group Analysis on Neuroimaging Data in Non-Euclidean Spaces

In Chapter 3 and 4, we addressed the need for more sensitive method to perform group analysis using neuroimaging data defined on graphs (i.e., cortical thickness measures on brain meshes or strength measures on brain connectivity). This becomes especially challenging when the given dataset is small sample-sized and the effect size is subtle. To resolve this problem, we developed a framework whose main contributions are

i) we derived a highly sensitive "multi-resolution" shape descriptor for performing group analysis in a population of subjects on signals defined on surfaces/shapes or signals defined on brain connectivity.

ii) we demonstrated the utility of the frameworks on various independent datasets and showed significant performance improvements over the standard baseline. In some cases, our framework is able to characterize changes in preclinical stages by comparing asymptomatic individuals with and without risk factors for AD. These experiments give strong evidence that a large number of neuroimaging data in non-Euclidean spaces can immediately benefit from these ideas with negligible additional cost.

8.1.2 Statistical Analysis of Images with Systematic Variations

In Chapter 5 we studied the problem of systematic variations in longitudinal imaging studies or multi-site imaging studies. The main contribution of the work introduced in Chapter 5 was to formalize this idea for immunity to the systematic variations in statistical analysis of imaging data, based on a new method in the harmonic analysis literature by Coifman and Hirn (Coifman and Hirn, 2014). In particular,

i) we derived operators from each image using the recent work in Diffusion Maps (Coifman and Lafon, 2006; Coifman and Hirn, 2014) and compared the spectrum of the operators to achieve differences between the images which are invariant to systematical variation.

ii) we described how the lower dimensional mapping obtained by the operators relate to a wavelet transform in non-Euclidean spaces.

iii) we provided experimental evidence in that the method facilitates statistical analysis of Pittsburgh compound B PET (PIB-PET) images and offers improvements over standard normalization methods used in the community.

8.1.3 Statistical Analysis of Images with Imperfect Registration

The work we introduced in Chapter 6 was to take a step towards registration error invariant analysis of brain imaging data. The contributions of our proposed algorithm were:

i) we leveraged recent results from harmonic analysis, namely, scattering coefficients to derive image representations that are provably invariant to

local deformations. This so-called scatter transform is obtained via a cascade of operations, involving wavelet expansions using an orthonormal basis derived from a function of the image. The downstream statistical questions can then be simply reformulated in terms of such representations, which are immune to nominal levels of errors in the given registration.

ii) we provided simulation results as well as empirical evidence obtained from experiments on real brain images showing that the image representation derived from our framework can provide more statistically significant and meaningful results in statistical brain image analyses.

8.1.4 Latent Variable Graphical Model Selection

In Chapter 7, we introduced the problem of identifying relationships between different variables which corresponds to different neural fiber pathways in the brain and covariates, the main bottleneck was the existence of latent variables that affects our observations. We proposed a novel method to resolve this problem using multi-resolution approach via harmonic analysis in non-Euclidean spaces, and its main contributions were

i) we demonstrated how this latent graphical model estimation problem can be viewed via the lens of harmonic analysis.

ii) we showed that by operating on the inverse covariance matrix via its associated graph (actually a wavelet transform of this graph), it becomes an inference problem expressed *in the frequency space*. The actual optimization requires no sophisticated solvers, we only need to perform a simple gradient descent on one variable that controls the band-pass filtering property of wavelets.

iii) Our motivating application was the analysis of the Human Connectome Project (HCP) dataset which includes more than ~350 covariates (and therefore, many latent variables) together with a rich set of imaging data. Here, we obtained neuroscientifically meaningful sparse models relating image-derived brain connectivity to covariates where alternative approaches yield uninterpretable results.

8.2 Future Work with Multi-resolution in Non-Euclidean Space

There exists a rich body of work in computer vision, machine learning and signal processing that demonstrates the power of a multi-resolution framework (Shen and Ip, 1999; Wink and Roerdink, 2004). The intuition here is that instead of looking at an object/image at a fixed distance, looking at it from different distances (scales) better captures its various properties. However, as the data that we collect in the real world are in *high dimensions* and exist in a *complex space*, traditional methods often become ill suited to analyze such data types. Therefore, we are interested in broadening the multi-resolution framework in non-Euclidean space for such complex data, not only for neuroimages, but also for general data that we cannot explore with traditional methods. Such a technique shares commonalities with many other machine learning and computer vision ideas, where the community is interested in analyzing data in low-dimensional spaces in which tasks become easier, in some sense.

8.2.1 Human Brain Connectome Analysis

The National Institute of Health (NIH) has recently launched the Human Connectome Project (HCP) (`http://www.humanconnectome.org`), releasing a large dataset (with imaging and non-imaging data) of 897 participants. The objective of the project is to identify variations in neural pathways that affect brain function and behavior. We plan to use ideas of multi-resolution in non-Euclidean space to analyze brain network data as in the HCP dataset, as a neural connectivity of the brain is naturally represented as a weighted graph in non-Euclidean space. Using the multi-resolution methodologies that we have developed, we plan to investigate new scientific hypothesis in the dataset, or validate the findings that have been reported using other imaging modalities. Further, we are interested in making use of local datasets that have been collected for their own purposes together with this HCP dataset. Those local data may be sufficient to answer a few questions for which the dataset was designed, but are statistically underpowered for analyses with other meaning-

ful factors which still may manifest strong effect. We can narrow down search spaces by first analyzing the HCP dataset, and use the result as a prior for the statistical analyses of other independent and smaller local datasets. Such a design will demonstrate improved statistical results with fewer number of hypothesis tests to compare. Combining the analyses of the HCP dataset and local datasets together, we will be able to answer many fundamental questions raised in neuroscience identifying which of the associations between different brain regions are closely related to clinically meaningful factors such as disease status, age, gender and genotype. These analyses will yield interesting scientific discoveries that have not been studied on structural relationships between brain regions and identify how the human brains function.

8.2.2 Analysis of Various Neurological Disorders / Cognitive Ability

We are interested in broadening the area of study to various neurological/psychiatric disorders such as Parkinson's disease, schizophrenia, bipolar symptoms, traumatic brain injury (TBI) and autism, and work on methods that will facilitate development of diagnosis techniques and new treatments. Analyzing various brain disorders is critical, since many disease specific symptoms in the human brain are shared across different disorders. Using multi-resolution analysis, we plan to study both structural and functional aspects of the brain to better understand different brain disorders, which will help increase the accuracy of diagnosis and narrow focuses for treatments. Understanding which specific structural/functional variations in the brain are closely related to cognitive decline, verbal inability and abnormal behaviors will give us a feasible interpretation of how the human brain functions. Such objectives are central to global large scale initiatives.

8.2.3 Adaptive Signal Recovery in Non-Euclidean Spaces

We are studying ideas related to multi-resolution analysis for estimating signals defined on graph nodes. The problem definition and technical details here are similar to traditional Matrix Completion problem (Candes and Recht, 2009;

Candes and Plan, 2010) where one is given with partial observations on some of the elements in a matrix and tries to recover the full matrix. When this regime is defined in a graph setting (Puy et al., 2016), given a graph and partial signals on the graph nodes, the objective is to estimate the full signal to "complete" a graph utilizing the graph structure and harmonic analysis. We are studying this problem from the perspective of multi-resolution and collaborative filtering, proposing an adaptive sampling scheme and estimation of the full signal by taking advantage of the bandlimited nature of signals. Recently, there are a number of data available that are represented as a graph such as social network, brain network, genome data, 3D mesh and so on, where data (i.e., signals) are collected on the nodes of the graphs. Our work here will broaden the goal of experimental setups required for the analysis of such complex data and advance scientific findings with less cost. We will exploit the properties of these graphs to provide important practical and immediate ramifications for many experimental design considerations in various scientific domains (Kim et al., 2016a; Hwa Kim et al., 2017).

REFERENCES

Adluru, Nagesh, Daniel J Destiche, Sharon Yuan-Fu Lu, Samuel T Doran, Alex C Birdsill, Kelsey E Melah, Ozioma C Okonkwo, Andrew L Alexander, N Maritza Dowling, Sterling C Johnson, et al. 2014. White matter microstructure in late middle-age: Effects of apolipoprotein E4 and parental family history of Alzheimer's disease. *NeuroImage: Clinical* 4:730–742.

Adluru, Nagesh, Hui Zhang, Andrew S Fox, et al. 2012. A diffusion tensor brain template for rhesus macaques. *Neuroimage* 59(1):306–318.

Akil, Huda, Maryann E Martone, and David C Van Essen. 2011. Challenges and opportunities in mining neuroscience data. *Science (New York, NY)* 331(6018):708.

Alexander, Daniel C, and Gareth J Barker. 2005. Optimal imaging parameters for fiber-orientation estimation in diffusion MRI. *Neuroimage* 27(2):357–367.

Antoine, J-P, L Demanet, L Jacques, and P Vandergheynst. 2002. Wavelets on the sphere: implementation and approximations. *Applied and Computational Harmonic Analysis* 13(3):177–200.

Antonini, Marc, Michel Barlaud, Pierre Mathieu, and Ingrid Daubechies. 1992. Image coding using wavelet transform. *IEEE Transactions on image processing* 1(2):205–220.

Ashburner, J, and K Friston. 1997. Multimodal image coregistration and partitioningâŁ"a unified framework. *Neuroimage* 6(3):209–217.

Ashburner, John, and Karl J Friston. 2000. Voxel-based morphometryâŁ"the methods. *Neuroimage* 11(6):805–821.

Aubry, Mathieu, Ulrich Schlickewei, and Daniel Cremers. 2011. The wave kernel signature: A quantum mechanical approach to shape analysis. In *Iccv workshops*, 1626–1633. IEEE.

Banerjee, Onureena, Laurent El Ghaoui, and Alexandre d'Aspremont. 2008. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *JMLR* 9:485–516.

Banerjee, Onureena, Laurent El Ghaoui, Alexandre d'Aspremont, and Georges Natsoulis. 2006. Convex optimization techniques for fitting sparse Gaussian graphical models. In *Icml*, 89–96. ACM.

Basser, Peter J, Sinisa Pajevic, Carlo Pierpaoli, Jeffrey Duda, and Akram Aldroubi. 2000. In vivo fiber tractography using dt-mri data. *Magnetic resonance in medicine* 44(4):625–632.

Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. 2006. Surf: Speeded up robust features. *Computer vision–ECCV 2006* 404–417.

Bendlin, Barbara B, Michele L Ries, Elisa Canu, Aparna Sodhi, Mariana Lazar, Andrew L Alexander, Cynthia M Carlsson, Mark A Sager, Sanjay Asthana, and Sterling C Johnson. 2010. White matter is altered with parental family history of Alzheimer's disease. *Alzheimer's & dementia* 6(5):394–403.

Benjamini, Yoav, and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57(1):pp. 289–300.

Booth, Tom, Mark E Bastin, Lars Penke, et al. 2013. Brain white matter tract integrity and cognitive abilities in community-dwelling older people: the Lothian birth cohort, 1936. *Neuropsych.* 27(5):595.

Braak, Heiko, and Eva Braak. 1995. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of aging* 16(3):271–278.

Bracewell, Ronald Newbold, and Ronald N Bracewell. 1986. *The Fourier transform and its applications*, vol. 31999. McGraw-Hill New York.

Brier, Matthew, Jewell Thomas, and Beau Ances. 2014a. Network dysfunction in alzheimerâĽ™s disease: Refining the disconnection hypothesis. *Brain connectivity* 4(5):299–311.

Brier, Matthew R, Jewell B Thomas, Anne M Fagan, Jason Hassenstab, David M Holtzman, Tammie L Benzinger, John C Morris, and Beau M Ances. 2014b. Functional connectivity and graph theory in preclinical alzheimer's disease. *Neurobiology of aging* 35(4):757–768.

Brislawn, Christopher. 1991. Traceable integral kernels on countably generated measure spaces. *Pacific Journal of Mathematics* 150(2):229–240.

Bronstein, Michael M, and Iasonas Kokkinos. 2010. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Cvpr*, 1704–1711. IEEE.

Bruna, Joan, and Stéphane Mallat. 2013. Invariant scattering convolution networks. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35(8): 1872–1886.

Buckner, Randy L, Abraham Z Snyder, Benjamin J Shannon, Gina LaRossa, Rimmon Sachs, Anthony F Fotenos, Yvette I Sheline, William E Klunk, Chester A Mathis, John C Morris, et al. 2005. Molecular, structural, and functional characterization of alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *The Journal of Neuroscience* 25(34):7709–7717.

Bullmore, Ed, and Olaf Sporns. 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews. Neuroscience* 10(3):186.

Canales-Rodríguez, Erick Jorge, Joaquim Radua, Edith Pomarol-Clotet, Salvador Sarró, Yasser Alemán-Gómez, Yasser Iturria-Medina, and Raymond Salvador. 2013. Statistical analysis of brain tissue images in the wavelet domain: Wavelet-based morphometry. *NeuroImage* 72:214–226.

Candes, Emmanuel J, and Yaniv Plan. 2010. Matrix completion with noise. *Proceedings of the IEEE* 98(6):925–936.

Candes, Emmanuel J, and Benjamin Recht. 2009. Exact matrix completion via convex optimization. *Foundations of Computational mathematics* 9(6):717–772.

Castellanos, F Xavier, Patti P Lee, Wendy Sharp, Neal O Jeffries, Deanna K Greenstein, Liv S Clasen, Jonathan D Blumenthal, Regina S James, Christen L Ebens, James M Walter, et al. 2002. Developmental trajectories of brain volume abnormalities in children and adolescents with attention-deficit/hyperactivity disorder. *Jama* 288(14):1740–1748.

Chan, Dennis, Nick C Fox, Rachael I Scahill, et al. 2001. Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Annals of Neurology* 49(4):433–442.

Chandrasekaran, Venkat, Pablo A Parrilo, Alan S Willsky, et al. 2012. Latent variable graphical model selection via convex optimization. *The Annals of Statistics* 40(4):1935–1967.

Chang, S Grace, Bin Yu, and Martin Vetterli. 2000. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on image processing* 9(9):1532–1546.

Chao, Samuel T, John H Suh, Shanker Raja, Shih-Yuan Lee, and Gene Barnett. 2001. The sensitivity and specificity of FDG PET in distinguishing recurrent brain tumor from radionecrosis in patients treated with stereotactic radiosurgery. *International journal of cancer* 96(3):191–197.

Chetelat, G, B Landeau, F Eustache, F Mezenge, F Viader, V De La Sayette, B Desgranges, and J-C Baron. 2005. Using voxel-based morphometry to map the structural changes associated with rapid conversion in mci: a longitudinal mri study. *Neuroimage* 27(4):934–946.

Chételat, Gaël, Renaud La Joie, Nicolas Villain, Audrey Perrotin, Vincent de La Sayette, Francis Eustache, and Rik Vandenberghe. 2013. Amyloid imaging in cognitively normal individuals, at-risk populations and preclinical Alzheimer's disease. *NeuroImage: Clinical* 2:356–365.

Cho, Youngsang, Joon-Kyung Seong, Yong Jeong, and Sung Yong Shin. 2012. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *Neuroimage* 59(3):2217–2230.

Chung, M.K. 2006. Heat kernel smoothing on unit sphere. In *Biomed. imaging: Nano to macro, 2006. 3rd ieee int. symposium on*, 992 –995.

Chung, M.K., K.M. Dalton, and R.J. Davidson. 2008. Tensor-based cortical surface morphometry via weighted spherical harmonic representation. *Med. Imaging, IEEE Trans. on* 27(8):1143 –1151.

Chung, M.K., K.M. Dalton, S. Li, et al. 2007. Weighted Fourier series representation and its application to quantifying the amount of gray matter. *Med. Imaging, IEEE Trans. on* 26(4):566 –581.

Chung, M.K., S.M. Robbins, K.M. Dalton, et al. 2005. Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage* 25(4):1256 – 1265.

Coifman, Ronald R, and Matthew J Hirn. 2014. Diffusion maps for changing data. *Applied and Computational Harmonic Analysis* 36(1):79–107.

Coifman, Ronald R, and Stéphane Lafon. 2006. Diffusion maps. *Applied and Computational Harmonic Analysis* 21(1):5–30.

Coifman, R.R., and M. Maggioni. 2006. Diffusion wavelets. *Applied and Computational Harmonic Analysis* 21(1):53 – 94.

Colcombe, Stanley J, Kirk I Erickson, Naftali Raz, Andrew G Webb, Neal J Cohen, Edward McAuley, and Arthur F Kramer. 2003. Aerobic fitness reduces brain tissue loss in aging humans. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 58(2):M176–M180.

Cook, PA, Y Bai, SKKS Nedjati-Gilani, KK Seunarine, MG Hall, GJ Parker, and DC Alexander. 2006. Camino: open-source diffusion-MRI reconstruction and processing. In *14th scientific meeting of the international society for magnetic resonance in medicine*, vol. 2759. Seattle WA, USA.

Corder, EH, AM Saunders, WJ Strittmatter, et al. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261(5123):921–923.

Courchesne, Eric, CM Karns, HR Davis, R Ziccardi, RA Carper, ZD Tigue, HJ Chisum, P Moses, K Pierce, C Lord, et al. 2001. Unusual brain growth patterns in early life in patients with autistic disorder an MRI study. *Neurology* 57(2):245–254.

Craik, Fergus IM, Tara M Moroz, Morris Moscovitch, Donald T Stuss, Gordon Winocur, Endel Tulving, and Shitij Kapur. 1999. In search of the self: A positron emission tomography study. *Psychological Science* 10(1):26–34.

Crossley, Nicolas A, Andrea Mechelli, Jessica Scott, Francesco Carletti, Peter T Fox, Philip McGuire, and Edward T Bullmore. 2014. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* 137(8):2382–2395.

Dahl, David B, and Michael A Newton. 2007. Multiple hypothesis testing by clustering treatment effects. *Journal of the American Statistical Association* 102(478):517–526.

Damoiseaux, Jessica S, Katherine E Prater, Bruce L Miller, and Michael D Greicius. 2012. Functional connectivity tracks clinical deterioration in Alzheimer's disease. *Neurobiology of aging* 33(4):828–e19.

d'Aspremont, Alexandre, Onureena Banerjee, and Laurent El Ghaoui. 2008. First-order methods for sparse covariance selection. *SIAM Journal on Matrix Analysis and Applications* 30(1):56–66.

Daubechies, Ingrid. 1990. The wavelet transform, time-frequency localization and signal analysis. *Information Theory, IEEE Transactions on* 36(5):961–1005.

Davis, Jason V, Brian Kulis, Prateek Jain, et al. 2007. Information-theoretic metric learning. In *Icml*, 209–216. ACM.

Davis, Simon W, Nancy A Dennis, Norbou G Buchler, et al. 2009. Assessing the effects of age on long white matter tracts using diffusion tensor tractography. *Neuroimage* 46(2):530–541.

Davison, Elizabeth N, Kimberly J Schlesinger, Danielle S Bassett, Mary-Ellen Lynall, Michael B Miller, Scott T Grafton, and Jean M Carlson. 2015. Brain network adaptability across task states. *PLoS computational biology* 11(1): e1004029.

DeKosky, Steven T, and Stephen W Scheff. 1990. Synapse loss in frontal cortex biopsies in Alzheimer's disease: correlation with cognitive severity. *Annals of neurology* 27(5):457–464.

Delbeuck, Xavier, Martial Van der Linden, and Fabienne Collette. 2003. Alzheimer's disease as a disconnection syndrome? *Neuropsychology review* 13(2):79–92.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *JRSS-B* 1–38.

Dennis, Nancy A, and Roberto Cabeza. 2008. Neuroimaging of healthy cognitive aging. *The handbook of aging and cognition* 3:1–54.

Dickerson, Bradford C, Eric Feczko, Jean C Augustinack, Jenni Pacheco, John C Morris, Bruce Fischl, and Randy L Buckner. 2009. Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiology of aging* 30(3):432–440.

Dickerson, Bradford C, and David A Wolk. 2012. MRI cortical thickness biomarker predicts AD-like CSF and cognitive decline in normal adults. *Neurology* 78(2):84–90.

Ding, Bei, Hua-wei Ling, Yong Zhang, Juan Huang, Huan Zhang, Tao Wang, and Fu Hua Yan. 2014. Pattern of cerebral hyperperfusion in alzheimerâŁ™s disease and amnestic mild cognitive impairment using voxel-based analysis of 3D arterial spin-labeling imaging: initial experience. *Clinical interventions in aging* 9:493.

Diniz, Breno SO, Jony A Pinto Jr, and Orestes Vicente Forlenza. 2008. Do CSF total tau, phosphorylated tau, and β-amyloid 42 help to predict progression of mild cognitive impairment to Alzheimer's disease? a systematic review

and meta-analysis of the literature. *The World Journal of Biological Psychiatry* 9(3):172–182.

Donahue, Jeff, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. Decaf: A deep convolutional activation feature for generic visual recognition. In *Icml*, vol. 32, 647–655.

Driemeyer, Joenna, Janina Boyke, Christian Gaser, Christian Büchel, and Arne May. 2008. Changes in gray matter induced by learningâĽ"revisited. *PLoS One* 3(7):e2669.

Dunnett, Charles W. 1955. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* 50(272):1096–1121.

Eismann, Michael T, Joseph Meola, and Russell C Hardie. 2008. Hyperspectral change detection in the presenceof diurnal and seasonal variations. *IEEE Geoscience and Remote Sensing* 46(1):237–249.

Elidan, Gal, Iftach Nachman, and Nir Friedman. 2007. Ideal parent structure learning for continuous variable Bayesian networks. *JMLR*.

Erkinjuntti, Timo, Leena Ketonen, Raimo Sulkava, Jorma Sipponen, MATTI Vuorialho, and MATTI Iivanainen. 1987. Do white matter changes on MRI and CT differentiate vascular dementia from Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 50(1):37–42.

Fagan, Anne M, Catherine M Roe, Chengjie Xiong, Mark A Mintun, John C Morris, and David M Holtzman. 2007. Cerebrospinal fluid tau/β-amyloid42 ratio as a prediction of cognitive decline in nondemented older adults. *Archives of neurology* 64(3):343–349.

Fang, Yi, Mengtian Sun, Minhyong Kim, and Karthik Ramani. 2011. Heat-mapping: A robust approach toward perceptually consistent mesh segmentation. In *Cvpr*, 2145–2152. IEEE.

Filippi, Massimo, and Federica Agosta. 2011. Structural and functional network connectivity breakdown in Alzheimer's disease studied with magnetic resonance imaging techniques. *Journal of Alzheimer's Disease* 24(3):455–474.

Fingelkurts, Andrew A, Alexander A Fingelkurts, Sergio Bagnato, Cristina Boccagni, and Giuseppe Galardi. 2012. Dmn operational synchrony relates to self-consciousness: evidence from patients in vegetative and minimally conscious states. *The open neuroimaging journal* 6:55.

Fischl, Bruce. 2012. Freesurfer. *NeuroImage* 62(2):774–781.

Fleisher, Adam S, Ayesha Sherzai, Curtis Taylor, Jessica BS Langbaum, Kewei Chen, and Richard B Buxton. 2009. Resting-state bold networks versus task-associated functional mri for distinguishing alzheimer's disease risk groups. *Neuroimage* 47(4):1678–1690.

Fornito, Alex, Andrew Zalesky, and Michael Breakspear. 2013. Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage* 80: 426–444.

Freeden, Willi, and Ulrich Windheuser. 1996. Spherical wavelet transform and its discretization. *Advances in Computational Mathematics* 5(1):51–94.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2008. Sparse inverse covariance estimation with the graphical LASSO. *Biostatistics* 9(3):432–441.

Glasser, Matthew F, Stamatios N Sotiropoulos, J Anthony Wilson, et al. 2013. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80:105–124.

Glenn, Orit A, Roland G Henry, Jeffrey I Berman, Patrick C Chang, Steven P Miller, Daniel B Vigneron, and A James Barkovich. 2003. Dti-based three-dimensional tractography detects differences in the pyramidal tracts of infants and children with congenital hemiparesis. *Journal of Magnetic Resonance Imaging* 18(6):641–648.

Gómez-Isla, Teresa, Richard Hollister, Howard West, Stina Mui, John H Growdon, Ronald C Petersen, Joseph E Parisi, and Bradley T Hyman. 1997. Neuronal loss correlates with but exceeds neurofibrillary tangles in alzheimer's disease. *Annals of neurology* 41(1):17–24.

Greicius, Michael D, Gaurav Srivastava, Allan L Reiss, and Vinod Menon. 2004. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional mri. *Proceedings of the National Academy of Sciences of the United States of America* 101(13):4637–4642.

Greicius, Michael D, Kaustubh Supekar, Vinod Menon, and Robert F Dougherty. 2009. Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral cortex* 19(1):72–78.

Gu, Lie, Eric P Xing, and Takeo Kanade. 2007. Learning GMRF structures for spatial priors. In *Cvpr*, 1–6.

Gusnard, Debra A, Erbil Akbudak, Gordon L Shulman, and Marcus E Raichle. 2001. Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proceedings of the National Academy of Sciences* 98(7):4259–4264.

Haacke, E Mark, Robert W Brown, Michael R Thompson, Ramesh Venkatesan, et al. 1999. *Magnetic resonance imaging: physical principles and sequence design*, vol. 82. Wiley-Liss New York:.

Hahn, Klaus, Nicholas Myers, Sergei Prigarin, Karsten Rodenacker, Alexander Kurz, Hans Förstl, Claus Zimmer, Afra M Wohlschläger, and Christian Sorg. 2013. Selectively and progressively disrupted structural connectivity of functional brain networks in Alzheimer's diseaseâŁ"revealed by a novel framework to analyze edge distributions of networks detecting disruptions with strong statistical evidence. *NeuroImage* 81:96–109.

Hammond, D.K., P. Vandergheynst, and R. Gribonval. 2011. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis* 30(2):129 – 150.

Harary, F. 1969. *Graph theory*. Addison-Wesley, Reading, MA.

Harris, Fredric J. 1978. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE* 66(1):51–83.

Haykin, Simon, and Barry Van Veen. 2007. *Signals and systems*. John Wiley & Sons.

Herrick, Rick, Michael McKay, Timothy Olsen, et al. 2014. Data dictionary services in XNAT and the Human Connectome Project. *Frontiers in neuroinformatics* 8.

van den Heuvel, Martijn, René Mandl, Judith Luigjes, and Hilleke Hulshoff Pol. 2008. Microstructural organization of the cingulum tract and the level of default mode functional connectivity. *The Journal of neuroscience* 28(43): 10844–10851.

Hinrichs, Chris, Vikas Singh, Lopamudra Mukherjee, Guofan Xu, Moo K Chung, and Sterling C Johnson. 2009. Spatially augmented lpboosting for ad classification with evaluations on the adni dataset. *NeuroImage* 48(1):138–149.

Hirschmuller, Heiko. 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Cvpr*, vol. 2, 807–814. IEEE.

———. 2008. Stereo processing by semiglobal matching and mutual information. *IEEE PAMI* 30(2):328–341.

Hodge, Michael R, William Horton, Timothy Brown, et al. 2015. ConnectomeDB-sharing human brain connectivity data. *NeuroImage*.

Hodsman, AB, M Kisiel, JD Adachi, LJ Fraher, and PH Watson. 2000. Histomorphometric evidence for increased bone turnover without change in cortical thickness or porosity after 2 years of cyclical hpth (1-34) therapy in women with severe osteoporosis. *Bone* 27(2):311–318.

Honorio, Jean, and Dimitris Samaras. 2010. Multi-task learning of Gaussian graphical models. In *Icml*, 447–454.

Hou, Tingbo, and Hong Qin. 2013. Admissible diffusion wavelets and their applications in space-frequency processing. *Visualization and Computer Graphics, IEEE Trans. on* 19(1):3–15.

Hsu, Jason. 1996. *Multiple comparisons: theory and methods*. CRC Press.

Huang, Shuai, Jing Li, Liang Sun, et al. 2010. Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation. *NeuroImage* 50(3):935–949.

Huang, Wenyong, Chengxuan Qiu, Eva von Strauss, Bengt Winblad, and Laura Fratiglioni. 2004. APOE genotype, family history of dementia, and Alzheimer disease risk: a 6-year follow-up study. *Archives of neurology* 61(12): 1930–1934.

Hull, Alastair M. 2002. Neuroimaging findings in post-traumatic stress disorder. *The British Journal of Psychiatry* 181(2):102–110.

Hwa Kim, Won, Mona Jalal, Seongjae Hwang, Sterling C. Johnson, and Vikas Singh. 2017. Online graph completion: Multivariate signal recovery in computer vision. In *Cvpr*.

Hyde, Krista L, Jason P Lerch, Robert J Zatorre, Timothy D Griffiths, Alan C Evans, and Isabelle Peretz. 2007. Cortical thickness in congenital amusia: when less is better than more. *The Journal of Neuroscience* 27(47):13028–13032.

Iglesias, Juan Eugenio, Ender Konukoglu, Darko Zikic, et al. 2013. Is synthesizing MRI contrast useful for inter-modality analysis? In *Miccai*, 631–638. Springer.

Ikonomovic, Milos D, William E Klunk, Eric E Abrahamson, Chester A Mathis, Julie C Price, Nicholas D Tsopelas, Brian J Lopresti, Scott Ziolko, Wenzhu Bi, William R Paljug, et al. 2008. Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease. *Brain* 131(6):1630–1645.

Ittner, Lars M, and Jürgen Götz. 2010. Amyloid-β and tau – a toxic pas de deux in Alzheimer's disease. *Nature Reviews Neuroscience* 12(2):67–72.

Jack, Clifford R, David S Knopman, William J Jagust, Ronald C Petersen, Michael W Weiner, Paul S Aisen, Leslie M Shaw, Prashanthi Vemuri, Heather J Wiste, Stephen D Weigand, et al. 2013. Update on hypothetical model of AlzheimerâŁ™s disease biomarkers. *Lancet neurology* 12(2):207.

Jack, CR, MM Shiung, JL Gunter, PC OâŁ™brien, SD Weigand, David S Knopman, Bradley F Boeve, Robert J Ivnik, Glenn E Smith, RH Cha, et al. 2004. Comparison of different mri brain atrophy rate measures with clinical disease progression in ad. *Neurology* 62(4):591–600.

Jack Jr, Clifford R, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology* 9(1):119–128.

Jagust, William, Amy Gitcho, Felice Sun, Beth Kuczynski, Dan Mungas, and Mary Haan. 2006. Brain imaging evidence of preclinical alzheimer's disease in normal aging. *Annals of neurology* 59(4):673–681.

Jenkinson, Mark, and Stephen Smith. 2001. A global optimisation method for robust affine registration of brain images. *Medical image analysis* 5(2):143–156.

Johnson, Karl J. 2008. Magnetic resonance imaging. In *Imaging in pediatric skeletal trauma*, 59–77. Springer.

Johnson, SC, TW Schmitz, CH Moritz, ME Meyerand, HA Rowley, AL Alexander, KW Hansen, CE Gleason, CM Carlsson, ML Ries, et al. 2006. Activation of brain regions vulnerable to alzheimer's disease: the effect of mild cognitive impairment. *Neurobiology of aging* 27(11):1604–1612.

Johnson, Sterling C, Michele L Ries, Timothy M Hess, Cynthia M Carlsson, Carey E Gleason, Andrew L Alexander, Howard A Rowley, Sanjay Asthana, and Mark A Sager. 2007. Effect of Alzheimer disease risk on brain function during self-appraisal in healthy middle-aged adults. *Archives of general psychiatry* 64(10):1163–1171.

Johnson, Sterling C., Asenath La Rue, Bruce P. Hermann, Guofan Xu, Rebecca L. Koscik, Erin M. Jonaitis, Barbara B. Bendlin, Kirk J. Hogan, Allen D. Roses, Ann M. Saunders, Michael W. Lutz, Sanjay Asthana, Robert C. Green, and Mark A. Sager. 2011. The effect of TOMM40 poly-T length on gray matter volume and cognition in middle-aged persons with APOE $\epsilon 3/\epsilon 3$ genotype. *Alzheimer's and Dementia* 7(4):456 – 465.

Jones, Derek K, and Alexander Leemans. 2011. Diffusion tensor imaging. *Magnetic Resonance Neuroimaging: Methods and Protocols* 127–144.

Jordan, Michael I, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. 1999. An introduction to variational methods for graphical models. *Machine learning* 37(2):183–233.

Jovicich, Jorge, Silvester Czanner, Douglas Greve, Elizabeth Haley, Andre van der Kouwe, Randy Gollub, David Kennedy, Franz Schmitt, Gregory Brown, James MacFall, et al. 2006. Reliability in multi-site structural mri studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage* 30(2):436–443.

Karlsgodt, Katherine H, Theo GM van Erp, Russell A Poldrack, et al. 2008. Diffusion tensor imaging of the superior longitudinal fasciculus and working memory in recent-onset schizophrenia. *Bio. psych.* 63(5):512–518.

Keppel, G, and TD Wickens. 2004. Simultaneous comparisons and the control of type i errors. *Design and analysis: A researcherâĽ™s handbook. 4th ed. Upper Saddle River (NJ): Pearson Prentice Hall. p* 111–130.

Kim, Junhwan, Vladimir Kolmogorov, and Ramin Zabih. 2003. Visual correspondence using energy minimization and mutual information. In *Iccv*, 1033–1040. IEEE.

Kim, Won Hwa, Nagesh Adluru, Moo K Chung, et al. 2013a. Multi-resolutional brain network filtering and analysis via wavelets on non-euclidean space. In *Miccai*, 643–651. Springer.

———. 2015a. Multi-resolution statistical analysis of brain connectivity graphs in preclinical Alzheimer's disease. *NeuroImage* 118:103–117.

Kim, Won Hwa, Moo K Chung, and Vikas Singh. 2013b. Multi-resolution shape analysis via Non-euclidean wavelets: Applications to mesh segmentation and surface alignment problems. In *Cvpr*, 2139–2146. IEEE.

Kim, Won Hwa, Seong Jae Hwang, Nagesh Adluru, et al. 2016a. Adaptive signal recovery on graphs via harmonic analysis for experimental design in neuroimaging. In *Eccv*, 188–205. Springer.

Kim, Won Hwa, Hyunwoo J Kim, Nagesh Adluru, et al. 2016b. Latent variable graphical model selection using harmonic analysis: Applications to the human connectome project (HCP). In *Cvpr*. IEEE.

Kim, Won Hwa, Deepti Pachauri, Charles Hatt, et al. 2012. Wavelet based multi-scale shape features on arbitrary surfaces for cortical thickness discrimination. In *Nips*, 1250–1258.

Kim, Won Hwa, Sathya N Ravi, Sterling C Johnson, Ozioma C Okonkwo, and Vikas Singh. 2015b. On statistical analysis of neuroimages with imperfect registration. In *Iccv*, 666–674.

Kim, Won Hwa, Vikas Singh, Moo K Chung, Nagesh Adluru, Barbara B Bendlin, and Sterling C Johnson. 2015c. Multi-resolution statistical analysis on graph structured data in neuroimaging. In *2015 ieee 12th international symposium on biomedical imaging (isbi)*, 1548–1551. IEEE.

Kim, Won Hwa, Vikas Singh, Moo K Chung, et al. 2014. Multi-resolutional shape features via non-Euclidean wavelets: Applications to statistical analysis of cortical thickness. *NeuroImage* 93:107–123.

Klein, Arno, Jesper Andersson, Babak A Ardekani, et al. 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage* 46(3):786–802.

Koch, Kathrin, Nicholas E Myers, Jens Göttler, Lorenzo Pasquini, Timo Grimmer, Stefan Förster, Andrei Manoliu, Julia Neitzel, Alexander Kurz, Hans Förstl, et al. 2014. Disrupted intrinsic networks link amyloid-β pathology and impaired cognition in prodromal alzheimer's disease. *Cerebral Cortex* bhu151.

Konrad, Kerstin, and Simon B Eickhoff. 2010. Is the ADHD brain wired differently? a review on structural and functional connectivity in attention deficit hyperactivity disorder. *Human brain mapping* 31(6):904–916.

Kowalski, Emmanuel. 2014. *An introduction to the representation theory of groups*, vol. 155. American Mathematical Society.

Kubicki, Marek, Carl-Fredrik Westin, Robert W McCarley, et al. 2005. The application of DTI to investigate white matter abnormalities in schizophrenia. *Ann. of the NY Academy of Sci.* 1064(1):134–148.

LaFerla, F. 2008. Amyloid-β and tau in AlzheimerâŁ™s diseaseâŁ™. *Nature Reviews Neuroscience*.

Lam, Clifford, and Jianqing Fan. 2009. Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. of stat.* 37(6B):4254.

Lassmann, H, P Fischer, and K Jellinger. 1993. Synaptic pathology of alzheimer's diseasea. *Annals of the New York Academy of Sciences* 695(1):59–64.

Lawrence, Steve, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. 1997. Face recognition: A convolutional neural-network approach. *Neural Networks, IEEE Transactions on* 8(1):98–113.

Lazar, Mariana, David M Weinstein, Jay S Tsuruda, Khader M Hasan, Konstantinos Arfanakis, M Elizabeth Meyerand, Benham Badie, Howard A Rowley, Victor Haughton, Aaron Field, et al. 2003. White matter tractography using diffusion tensor deflection. *Human brain mapping* 18(4):306–321.

Le Bihan, Denis, Jean-François Mangin, Cyril Poupon, Chris A Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat. 2001. Diffusion tensor imaging: concepts and applications. *Journal of magnetic resonance imaging* 13(4): 534–546.

Lehmann, Manja, Sebastian J Crutch, Gerard R Ridgway, Basil H Ridha, Josephine Barnes, Elizabeth K Warrington, Martin N Rossor, and Nick C Fox. 2011. Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease. *Neurobiology of aging* 32(8): 1466–1476.

Lemaitre, Herve, Aaron L Goldman, Fabio Sambataro, Beth A Verchinski, Andreas Meyer-Lindenberg, Daniel R Weinberger, and Venkata S Mattay. 2012. Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume. *Neurobiology of aging* 33(3):617–e1.

de Leon, Mony J, Ajax E George, Barry Reisberg, SH Ferris, Alan Kluger, Leonidas A Stylopoulos, JD Miller, Mary Ellen La Regina, Clara Chen, and Jacob Cohen. 1989. Alzheimer's disease: longitudinal CT studies of ventricular change. *American Journal of Roentgenology* 152(6):1257–1262.

Leonardi, N., and D. Van De Ville. 2013. Tight wavelet frames on multislice graphs. *Signal Processing, IEEE Transactions on* 61(13):3357–3367.

Leow, Alex D, Igor Yanovsky, Neelroop Parikshak, et al. 2009. Alzheimer's disease neuroimaging initiative: a one-year follow up study using tensor-based morphometry correlating degenerative rates, biomarkers and cognition. *Neuroimage* 45(3):645–655.

Lerch, Jason P, Jens Pruessner, Alex P Zijdenbos, D Louis Collins, Stefan J Teipel, Harald Hampel, and Alan C Evans. 2008. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiology of aging* 29(1):23–30.

Lerch, Jason P, Jens C Pruessner, Alex Zijdenbos, Harald Hampel, Stefan J Teipel, and Alan C Evans. 2005. Focal decline of cortical thickness in

Alzheimer's disease identified by computational neuroanatomy. *Cerebral cortex* 15(7):995–1001.

Li, Shi-Jiang, Zhu Li, Gaohong Wu, Mei-Jie Zhang, Malgorzata Franczak, and Piero G Antuono. 2002. Alzheimer disease: Evaluation of a functional mr imaging index as a marker 1. *Radiology* 225(1):253–259.

Lindeberg, Tony. 1993. *Scale-space theory in computer vision*. Springer.

———. 1994. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics* 21(1-2):225–270.

Ling, Haibin, and David W Jacobs. 2005. Deformation invariant image matching. In *Iccv*, vol. 2, 1466–1473. IEEE.

Liu, Han, Kathryn Roeder, and Larry Wasserman. 2010. Stability approach to regularization selection (stars) for high dimensional graphical models. In *Nips*, 1432–1440.

Lowe, David G. 2004. Distinctive image features from scale-invariant keypoints. *IJCV* 60(2):91–110.

Lu, Zhitao, Dong Youn Kim, and William A Pearlman. 2000. Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm. *IEEE transactions on Biomedical Engineering* 47(7):849–856.

Lucignani, Giovanni, Giovanni Paganelli, and Emilio Bombardieri. 2004. The use of standardized uptake values for assessing FDG uptake with PET in oncology: a clinical perspective. *Nuclear medicine communications* 25(7):651–656.

Maes, Frederik, Andre Collignon, Dirk Vandermeulen, et al. 1997. Multimodality image registration by maximization of mutual information. *IEEE TMI* 16(2):187–198.

Maintz, JB Antoine, and Max A Viergever. 1998. A survey of medical image registration. *Medical image analysis* 2(1):1–36.

Mallat, Stéphane. 2012. Group invariant scattering. *Communications on Pure and Applied Mathematics* 65(10):1331–1398.

Mallat, Stephane G. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 11(7):674–693.

Marcus, Daniel S, John Harwell, Timothy Olsen, et al. 2011. Informatics and data mining tools and strategies for the Human Connectome Project. *Frontiers in neuroinformatics* 5.

Marlin, Benjamin M, and Kevin P Murphy. 2009. Sparse Gaussian graphical models with unknown block structure. In *Icml*, 705–712.

Mattes, David, David R Haynor, Hubert Vesselle, et al. 2003. PET-CT image registration in the chest using free-form deformations. *IEEE TMI* 22(1):120–128.

McKhann, Guy, David Drachman, Marshall Folstein, Robert Katzman, Donald Price, and Emanuel M Stadlan. 1984. Clinical diagnosis of Alzheimer's disease report of the nincds-adrda work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology* 34(7):939–939.

McKhann, Guy M, David S Knopman, Howard Chertkow, Bradley T Hyman, Clifford R Jack Jr, Claudia H Kawas, William E Klunk, Walter J Koroshetz, Jennifer J Manly, Richard Mayeux, et al. 2011. The diagnosis of dementia due to AlzheimerâŁ™s disease: Recommendations from the national institute on aging-AlzheimerâŁ™s association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* 7(3):263–269.

Montine, Thomas J, Creighton H Phelps, Thomas G Beach, Eileen H Bigio, Nigel J Cairns, Dennis W Dickson, Charles Duyckaerts, Matthew P Frosch, Eliezer Masliah, Suzanne S Mirra, et al. 2012. National institute on aging–alzheimerâŁ™s association guidelines for the neuropathologic assessment of alzheimerâŁ™s disease: a practical approach. *Acta neuropathologica* 123(1):1–11.

Morel, Jean-Michel, and Guoshen Yu. 2009. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences* 2(2): 438–469.

Mosconi, L, R Mistur, R Switalski, M Brys, L Glodzik, K Rich, E Pirraglia, W Tsui, S De Santi, and MJ De Leon. 2009. Declining brain glucose metabolism in normal individuals with a maternal history of alzheimer disease. *Neurology* 72(6):513–520.

Mosconi, Lisa, John Murray, Wai H Tsui, Yi Li, Nicole Spector, Alexander Goldowsky, Schantel Williams, Ricardo Osorio, Pauline McHugh, Lidia Glodzik, et al. 2014. Brain imaging of cognitively normal individuals with 2 parents affected by late-onset AD. *Neurology* 82(9):752–760.

Narang, Sunil K, Yung Hsuan Chao, and Antonio Ortega. 2012. Graph-wavelet filterbanks for edge-aware image processing. In *Statistical signal processing workshop (ssp), 2012 ieee*, 141–144. IEEE.

Narang, Sunil K, and Antonio Ortega. 2012. Perfect reconstruction two-channel wavelet filter banks for graph structured data. *Signal Processing, IEEE Transactions on* 60(6):2786–2799.

Nelson, Peter T, Irina Alafuzoff, Eileen H Bigio, Constantin Bouras, Heiko Braak, Nigel J Cairns, Rudolph J Castellani, Barbara J Crain, Peter Davies, Kelly Del Tredici, et al. 2012. Correlation of alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *Journal of neuropathology and experimental neurology* 71(5):362.

Nestor, Paul G, Marek Kubicki, Kevin M Spencer, et al. 2007. Attentional networks and cingulum bundle in chronic schizophrenia. *Schizophrenia research* 90(1):308–315.

Newman, DL, Geoffrey Dougherty, A Al Obaid, and H Al Hajrasy. 1998. Limitations of clinical CT in assessing cortical thickness and density. *Physics in medicine and biology* 43(3):619.

Niogi, SN, P Mukherjee, J Ghajar, C Johnson, RA Kolster, R Sarkar, H Lee, M Meeker, RD Zimmerman, GT Manley, et al. 2008. Extent of microstructural white matter injury in postconcussive syndrome correlates with impaired cognitive reaction time: a 3T diffusion tensor imaging study of mild traumatic brain injury. *American Journal of Neuroradiology* 29(5):967–973.

Notay, Yvan. 2002. Combination of Jacobi–Davidson and conjugate gradients for the partial symmetric eigenproblem. *Numerical Linear Algebra with Applications* 9(1):21–44.

O'Donnell, Shannon, Michael D Noseworthy, Brian Levine, and Maureen Dennis. 2005. Cortical thickness of the frontopolar area in typically developing children and adolescents. *Neuroimage* 24(4):948–954.

Ogawa, Seiji, Tso-Ming Lee, Alan R Kay, and David W Tank. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences* 87(24):9868–9872.

Okello, A, J Koivunen, P Edison, HA Archer, FE Turkheimer, K ua Någren, R Bullock, Z Walker, A Kennedy, NC Fox, et al. 2009. Conversion of amyloid positive and negative MCI to AD over 3 years an 11c-PIB PET study. *Neurology* 73(10):754–760.

Okonkwo, Ozioma C, Guofan Xu, Jennifer M Oh, N Maritza Dowling, Cynthia M Carlsson, Catherine L Gallagher, Alex C Birdsill, Matthew Palotti, Whitney Wharton, Bruce P Hermann, et al. 2012. Cerebral blood flow is diminished in asymptomatic middle-aged adults with maternal history of alzheimer's disease. *Cerebral Cortex* bhs381.

Osman, Nael F, and Jerry L Prince. 2004. Regenerating MR tagged images using harmonic phase (harp) methods. *IEEE Biomedical Engineering* 51(8): 1428–1433.

Oztoprak, Figen, Jorge Nocedal, Steven Rennie, et al. 2012. Newton-like methods for sparse inverse covariance estimation. In *Nips*.

Pachauri, D., C. Hinrichs, M.K. Chung, et al. 2011. Topology-based kernels with application to inference problems in Alzheimer's disease. *Medical Imaging, IEEE Transactions on* 30(10):1760 –1770.

Padma, MV, S Said, M Jacobs, DR Hwang, K Dunigan, M Satter, B Christian, J Ruppert, T Bernstein, G Kraus, et al. 2003. Prediction of pathology and survival by FDG PET in gliomas. *Journal of neuro-oncology* 64(3):227–237.

Palaniyappan, Lena. 2010. Computing cortical surface measures in schizophrenia. *The British Journal of Psychiatry* 196(5):414–414.

Panizzon, Matthew S, Christine Fennema-Notestine, Lisa T Eyler, Terry L Jernigan, Elizabeth Prom-Wormley, Michael Neale, Kristen Jacobson, Michael J Lyons, Michael D Grant, Carol E Franz, et al. 2009. Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex* 19(11):2728–2735.

Penke, L, S Muñoz Maniega, ME Bastin, et al. 2012. Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular psychiatry* 17(10):1026–1030.

Perez-Nievas, Beatriz G, Thor D Stein, Hwan-Ching Tai, Oriol Dols-Icardo, Thomas C Scotton, Isabel Barroeta-Espar, Leticia Fernandez-Carballo, Estibaliz Lopez de Munain, Jesus Perez, Marta Marquie, et al. 2013. Dissecting phenotypic traits linked to human resilience to alzheimerâŁ™s pathology. *Brain* awt171.

Petersen, RC, JC Stevens, M Ganguli, EG Tangalos, JL Cummings, and ST DeKosky. 2001. Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review) report of the quality standards subcommittee of the american academy of neurology. *Neurology* 56(9): 1133–1142.

Pluim, Josien PW, JB Antoine Maintz, and Max A Viergever. 2003. Mutual-information-based registration of medical images: a survey. *IEEE TMI* 22(8): 986–1004.

Prevrhal, Sven, Klaus Engelke, and Willi A Kalender. 1999. Accuracy limits for the determination of cortical width and density: the influence of object size and CT imaging parameters. *Physics in medicine and biology* 44(3):751.

Puy, Gilles, Nicolas Tremblay, Rémi Gribonval, et al. 2016. Random sampling of bandlimited signals on graphs. *Applied and Computational Harmonic Analysis*.

Querbes, Olivier, Florent Aubry, Jérémie Pariente, Jean-Albert Lotterie, Jean-François Démonet, Véronique Duret, Michèle Puel, Isabelle Berry, Jean-Claude Fort, Pierre Celsis, et al. 2009. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain* 132(8):2036–2047.

Raskutti, Garvesh, Bin Yu, Martin J Wainwright, et al. 2008. Model selection in gaussian graphical models: High-dimensional consistency of\ boldmath ell_1-regularized mle. In *Nips*, 1329–1336.

Reiman, Eric M, Kewei Chen, Xiaofen Liu, et al. 2009. Fibrillar amyloid-β burden in cognitively normal people at 3 levels of genetic risk for Alzheimer's disease. *Proceedings of the National Academy of Sciences* 106(16):6820–6825.

Reuter, Martin, Herminia Diana Rosas, and Bruce Fischl. 2010. Highly accurate inverse consistent registration: A robust approach. *NeuroImage* 53(4): 1181–1196.

Rinne, Juha O, David J Brooks, Martin N Rossor, Nick C Fox, Roger Bullock, William E Klunk, Chester A Mathis, Kaj Blennow, Jerome Barakos, Aren A Okello, et al. 2010. 11 C-PiB PET assessment of change in fibrillar amyloid-β load in patients with Alzheimer's disease treated with bapineuzumab: a phase 2, double-blind, placebo-controlled, ascending-dose study. *The Lancet Neurology* 9(4):363–372.

Risacher, Shannon L, Andrew J Saykin, John D Wes, Li Shen, Hiram A Firpi, and Brenna C McDonald. 2009. Baseline mri predictors of conversion from mci to probable ad in the adni cohort. *Current Alzheimer Research* 6(4):347–361.

Roche, Alexis, Grégoire Malandain, Xavier Pennec, et al. 1998. The correlation ratio as a new similarity measure for multimodal image registration. In *Miccai*, 1115–1124. Springer.

Roy, Snehashis, Aaron Carass, Navid Shiee, et al. 2010. MR contrast synthesis for lesion segmentation. In *Biomedical imaging: From nano to macro*, 932–935. IEEE.

Ruttimann, Urs E, Michael Unser, Robert R Rawlings, Daniel Rio, Nick F Ramsey, Venkata S Mattay, Daniel W Hommer, Joseph A Frank, and Daniel R Weinberger. 1998. Statistical analysis of functional MRI data in the wavelet domain. *Medical Imaging, IEEE Transactions on* 17(2):142–154.

Ryan, Natalie S, Shiva Keihaninejad, Timothy J Shakespeare, Manja Lehmann, Sebastian J Crutch, Ian B Malone, John S Thornton, Laura Mancini, Harpreet Hyare, Tarek Yousry, et al. 2013. Magnetic resonance imaging evidence for presymptomatic change in thalamus and caudate in familial alzheimerâŁ™s disease. *Brain* 136(5):1399–1414.

Salat, David H, Randy L Buckner, Abraham Z Snyder, Douglas N Greve, Rahul SR Desikan, Evelina Busa, John C Morris, Anders M Dale, and Bruce Fischl. 2004. Thinning of the cerebral cortex in aging. *Cerebral cortex* 14(7): 721–730.

Sardy, Sylvain, Paul Tseng, and Andrew Bruce. 2001. Robust wavelet denoising. *IEEE Transactions on Signal Processing* 49(6):1146–1152.

Savitz, Jonathan, and Wayne C Drevets. 2009. Bipolar and major depressive disorder: neuroimaging the developmental-degenerative divide. *Neuroscience & Biobehavioral Reviews* 33(5):699–771.

Scheinberg, Katya, Shiqian Ma, and Donald Goldfarb. 2010. Sparse inverse covariance selection via alternating linearization methods. In *Nips*, 2101–2109.

Scheltens, PH, D Leys, F Barkhof, et al. 1992. Atrophy of medial temporal lobes on MRI in probable Alzheimer's disease and normal ageing: diagnostic

value and neuropsychological correlates. *Journal of Neurology, Neurosurgery & Psychiatry* 55(10):967–972.

Schmitz, Taylor W, and Sterling C Johnson. 2007. Relevance to self: a brief review and framework of neural systems underlying appraisal. *Neuroscience & Biobehavioral Reviews* 31(4):585–596.

Shao, Junming, Nicholas Myers, Qinli Yang, Jing Feng, Claudia Plant, Christian Böhm, Hans Förstl, Alexander Kurz, Claus Zimmer, Chun Meng, et al. 2012. Prediction of alzheimer's disease using individual structural connectivity networks. *Neurobiology of aging* 33(12):2756–2765.

Shattuck, David W, Stephanie R Sandor-Leahy, Kirt A Schaper, et al. 2001. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage* 13(5):856–876.

Shaw, Philip, Dede Greenstein, Jason Lerch, L Clasen, R Lenroot, N Gogtay, A Evans, J Rapoport, and J Giedd. 2006a. Intellectual ability and cortical development in children and adolescents. *Nature* 440(7084):676–679.

Shaw, Philip, Jason Lerch, Deanna Greenstein, Wendy Sharp, Liv Clasen, Alan Evans, Jay Giedd, F Xavier Castellanos, and Judith Rapoport. 2006b. Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry* 63(5):540.

Sheline, Yvette I, Marcus E Raichle, Abraham Z Snyder, John C Morris, Denise Head, Suzhi Wang, and Mark A Mintun. 2010. Amyloid plaques disrupt resting state default mode network connectivity in cognitively normal elderly. *Biological psychiatry* 67(6):584–587.

Shen, Dinggang, and Horace HS Ip. 1999. Discriminative wavelet shape descriptors for recognition of 2-d patterns. *Pattern recognition* 32(2):151–166.

Simpson, Ivor JA, Mark W Woolrich, Jesper LR Andersson, Adrian R Groves, Julia Schnabel, et al. 2013. Ensemble learning incorporating uncertain registration. *TMI* 32(4):748–756.

Simpson, Ivor JA, MarkW Woolrich, Adrian R Groves, and Julia A Schnabel. 2011. Longitudinal brain mri analysis with uncertain registration. In *Miccai*, 647–654. Springer.

Smith, Stephen M, Mark Jenkinson, Heidi Johansen-Berg, Daniel Rueckert, Thomas E Nichols, Clare E Mackay, Kate E Watkins, Olga Ciccarelli, M Zaheer Cader, Paul M Matthews, et al. 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 31(4):1487–1505.

Smith, Stephen M, Mark Jenkinson, Mark W Woolrich, et al. 2004. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23:S208–S219.

Smith, Stephen M, Karla L Miller, Salimi-Khorshidi, et al. 2011. Network modelling methods for fMRI. *Neuroimage* 54(2):875–891.

Sorg, Christian, Valentin Riedl, Mark Mühlau, Vince D Calhoun, Tom Eichele, Leonhard Läer, Alexander Drzezga, Hans Förstl, Alexander Kurz, Claus Zimmer, et al. 2007. Selective changes of resting-state networks in individuals at risk for alzheimer's disease. *Proceedings of the National Academy of Sciences* 104(47):18760–18765.

Sotiropoulos, Stamatios N, Saad Jbabdi, Junqian Xu, et al. 2013. Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *Neuroimage* 80:125–143.

Sowell, Elizabeth R, Bradley S Peterson, Eric Kan, Roger P Woods, June Yoshii, Ravi Bansal, Dongrong Xu, Hongtu Zhu, Paul M Thompson, and Arthur W Toga. 2007. Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. *Cerebral Cortex* 17(7): 1550–1560.

Sowell, Elizabeth R, Paul M Thompson, Christiana M Leonard, Suzanne E Welcome, Eric Kan, and Arthur W Toga. 2004. Longitudinal mapping of cortical thickness and brain growth in normal children. *The Journal of Neuroscience* 24(38):8223–8231.

Sparks, BF, SD Friedman, DW Shaw, EH Aylward, D Echelard, Alan A Artru, Kenneth R Maravilla, JN Giedd, J Munson, G Dawson, et al. 2002. Brain structural abnormalities in young children with autism spectrum disorder. *Neurology* 59(2):184–192.

Sperling, Reisa A, Peter S LaViolette, Kelly O'Keefe, Jacqueline O'Brien, Dorene M Rentz, Maija Pihlajamaki, Gad Marshall, Bradley T Hyman, Dennis J Selkoe, Trey Hedden, et al. 2009. Amyloid deposition is associated with impaired default network function in older persons without dementia. *Neuron* 63(2):178–188.

Spitzer, Frank. 2013. *Principles of random walk*, vol. 34. Springer Science & Business Media.

Sun, Mengtian, Yi Fang, and Karthik Ramani. 2012. Center-shift: an approach towards automatic robust mesh segmentation (arms). In *Cvpr*, 630–637. IEEE.

Supekar, Kaustubh, Vinod Menon, Daniel Rubin, Mark Musen, and Michael D Greicius. 2008. Network analysis of intrinsic functional brain connectivity in alzheimer's disease. *PLoS computational biology* 4(6):e1000100.

Takeda, Mitsuo, Hideki Ina, and Seiji Kobayashi. 1982. Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *JosA* 72(1):156–160.

Thompson, Paul, and Arthur W Toga. 1996. A surface-based technique for warping three-dimensional images of the brain. *IEEE transactions on medical imaging* 15(4):402–417.

Thompson, Paul M, Kiralee M Hayashi, Elizabeth R Sowell, Nitin Gogtay, Jay N Giedd, Judith L Rapoport, Greig I De Zubicaray, Andrew L Janke, Stephen E Rose, James Semple, et al. 2004. Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. *Neuroimage* 23(Suppl 1):S2–S18.

Thompson, Wesley K, Joachim Hallmayer, and Ruth O'Hara. 2011. Design considerations for characterizing psychiatric trajectories across the lifespan:

Application to effects of apoe-e4 on cerebral cortical thickness in Alzheimer's disease. *American Journal of Psychiatry* 168(9):894–903.

Uğurbil, Kamil, Junqian Xu, Edward J Auerbach, et al. 2013. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage* 80:80–104.

Uhlhaas, Peter J, and Wolf Singer. 2006. Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron* 52(1): 155–168.

Ullman, Jodie B, and Peter M Bentler. 2003. *Structural equation modeling*. Wiley Online Library.

Vaillancourt, DE, MB Spraker, J Prodoehl, I Abraham, DM Corcos, XJ Zhou, CL Comella, and DM Little. 2009. High-resolution diffusion tensor imaging in the substantia nigra of de novo parkinson disease. *Neurology* 72(16):1378–1384.

Van De Ville, Dimitri, Thierry Blu, and Michael Unser. 2004. Integrated wavelet processing and spatial statistical testing of fMRI data. *NeuroImage* 23(4):1472–1485.

Van Essen, David C, Stephen M Smith, Deanna M Barch, et al. 2013. The WU-Minn Human Connectome Project: an overview. *Neuroimage* 80:62–79.

Varentsova, Anna, Shengwei Zhang, and Konstantinos Arfanakis. 2014. Development of a high angular resolution diffusion imaging human brain template. *NeuroImage* 91:177–186.

Varoquaux, Gaël, Alexandre Gramfort, Jean-Baptiste Poline, et al. 2010. Brain covariance selection: better individual functional connectivity models using population prior. In *Nips*, 2334–2342.

Viola, Paul, and William M Wells III. 1997. Alignment by maximization of mutual information. *ICCV* 24(2):137–154.

Wang, Kun, Meng Liang, Liang Wang, Lixia Tian, Xinqing Zhang, Kuncheng Li, and Tianzi Jiang. 2007. Altered functional connectivity in early alzheimer's disease: A resting-state fmri study. *Human brain mapping* 28(10):967–978.

Watkins, Kate E, Faraneh Vargha-Khadem, John Ashburner, Richard E Passingham, Alan Connelly, Karl J Friston, Richard SJ Frackowiak, Mortimer Mishkin, and David G Gadian. 2002. Mri analysis of an inherited speech and language disorder: structural brain abnormalities. *Brain* 125(3):465–478.

Welch, Peter. 1967. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics* 15(2):70–73.

Wells III, William M, Paul Viola, Hideki Atsumi, et al. 1996. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis* 1(1):35–51.

Whitwell, Jennifer L., Nirubol Tosakulwong, Stephen D. Weigand, Matthew L. Senjem, Val J. Lowe, Jeffrey L. Gunter, Bradley F. Boeve, David S. Knopman, Bradford C. Dickerson, Ronald C. Petersen, and Clifford R. Jack Jr. 2013. Does amyloid deposition produce a specific atrophic signature in cognitively normal subjects? *NeuroImage: Clinical* 2(0):249 – 257.

Williams, Leanne M, Andrew H Kemp, Kim Felmingham, Matthew Barton, Gloria Olivieri, Anthony Peduto, Evian Gordon, and Richard A Bryant. 2006. Trauma modulates amygdala and medial prefrontal responses to consciously attended fear. *Neuroimage* 29(2):347–357.

Wingfield, Arthur, and Murray Grossman. 2006. Language and the aging brain: patterns of neural compensation revealed by functional brain imaging. *Journal of Neurophysiology* 96(6):2830–2839.

Wink, Alle Meije, and Jos BTM Roerdink. 2004. Denoising functional MR images: a comparison of wavelet denoising and gaussian smoothing. *IEEE transactions on medical imaging* 23(3):374–387.

Wirth, Miranka, Cindee M Madison, Gil D Rabinovici, Hwamee Oh, Susan M Landau, and William J Jagust. 2013. Alzheimer's disease neurodegenerative biomarkers are associated with decreased cognitive function but not β-amyloid in cognitively normal older individuals. *The Journal of Neuroscience* 33(13):5553–5563.

Wolz, Robin, Valtteri Julkunen, Juha Koikkalainen, Eini Niskanen, Dong Ping Zhang, Daniel Rueckert, Hilkka Soininen, Jyrki Lötjönen, et al. 2011. Multimethod analysis of MRI images in early diagnostics of Alzheimer's disease. *PloS one* 6(10):e25446.

Wozniak, Jeffrey R, Linda Krach, Erin Ward, Bryon A Mueller, Ryan Muetzel, Sarah Schnoebelen, Andrew Kiragu, and Kelvin O Lim. 2007. Neurocognitive and neuroimaging correlates of pediatric traumatic brain injury: a diffusion tensor imaging (DTI) study. *Archives of Clinical Neuropsychology* 22(5):555–568.

Xu, Guofan, Donald G Mclaren, Michele L Ries, Michele E Fitzgerald, Barbara B Bendlin, Howard A Rowley, Mark A Sager, Craig Atwood, Sanjay Asthana, and Sterling C Johnson. 2009. The influence of parental history of Alzheimer's disease and apolipoprotein ε4 on the bold signal during recognition memory. *Brain* 132(2):383–391.

Yu, Peng, P Ellen Grant, Yuan Qi, Xiao Han, Florent Ségonne, Rudolph Pienaar, Evelina Busa, Jenni Pacheco, Nikos Makris, Randy L Buckner, et al. 2007. Cortical surface shape analysis based on spherical wavelets. *Medical Imaging, IEEE Transactions on* 26(4):582–597.

Yuan, Ming. 2010. High dimensional inverse covariance matrix estimation via linear programming. *JMLR* 11:2261–2286.

Yuan, Ming, and Yi Lin. 2007. Model selection and estimation in the Gaussian graphical model. *Biometrika* 94(1):19–35.

Zarei, Mojtaba, David Mataix-Cols, Isobel Heyman, et al. 2011. Changes in gray matter volume and white matter microstructure in adolescents with obsessive-compulsive disorder. *Biological psychiatry* 70(11):1083–1090.

Zhang, Daoqiang, Yaping Wang, Luping Zhou, et al. 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55(3):856–867.

Zhang, Hui, Brian B Avants, Paul A Yushkevich, John H Woo, Sumei Wang, Leo F McCluskey, Lauren B Elman, Elias R Melhem, and James C Gee. 2007a. High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. *Medical Imaging, IEEE Transactions on* 26(11):1585–1597.

Zhang, Hui, Paul A Yushkevich, Daniel C Alexander, and James C Gee. 2006a. Deformable registration of diffusion tensor mr images with explicit orientation optimization. *Medical image analysis* 10(5):764–785.

Zhang, Hui, Paul A Yushkevich, Daniel Rueckert, et al. 2007b. Unbiased white matter atlas construction using diffusion tensor images. In *Medical image computing and computer-assisted intervention–miccai 2007*, 211–218. Springer.

Zhang, Zhongbo, Siliang Ma, and Xiao Han. 2006b. Multiscale feature extraction of finger-vein patterns based on curvelets and local interconnection structure neural network. In *Pattern recognition, 2006. icpr 2006. 18th international conference on*, vol. 4, 145–148. IEEE.

Zhou, Hao, Vamsi K Ithapu, Sathya Narayanan Ravi, Vikas Singh, Grace Wahba, and Sterling C Johnson. 2016. Hypothesis testing in unsupervised domain adaptation with applications in alzheimer's disease. In *Advances in neural information processing systems*, 2496–2504.

Zhou, Hao Henry, Yilin Zhang, Vamsi K Ithapu, Sterling C Johnson, Grace Wahba, and Vikas Singh. 2017. When can multi-site datasets be pooled for regression? hypothesis tests, ℓ_2-consistency and neuroscience applications. In *International conference on machine learning*, 4170–4179.