WEIGHTED FOURIER IMAGE ANALYSIS AND MODELING

By

Shubing Wang

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

(Statistics)

at the

UNIVERSITY OF WISCONSIN – MADISON

2008

Abstract

A novel systematic framework of medical image analysis, *weighted Fourier series* (WFS) analysis is introduced. WFS is a combination of Fourier series and heat kernel smoothing. WFS effectively reduces the Gibbs phenomenon, improves the signal to noise ratio, and increases normality of the estimated errors in the WFS-based generalized linear models.

To address the computational inefficiency in the least squares estimation of WFS, much faster but less accurate *iterative residual fitting* (IRF) method has been proposed. The proposed *adaptive iterative regression* (AIR) technique inherits the computational efficiency of IRF and improves accuracy of IRF. AIR partitions the function space into a set of subspaces, and performs an extra orthogonalization procedure to reduce the bias of IRF estimation.

For robust and accurate curvature estimation, we propose a new curve curvature calculation method. This method is independent of parametrization so that it can be applied to improve curve parametrization. Then a curvature-based non-linear curve registration is proposed. Surface curvatures are calculated analytically using the recurrence properties of the derivatives of Legendre polynomials. A new curvature-based surface alignment is proposed. It is equivalent to the affine alignment using coordinates of the surfaces, but is computationally more efficient.

Keywords: Autism, Eigenvalues and eigenfunctions, Fast Fourier transform, Fourier series, Fourier transform, Full width half maximum, Gaussian and mean curvature, Gradient vector flow snakes, Heat kernel, Hilbert space, Model selection, Nonlinear registration, Random field theory, Spherical harmonics, Spherical transform, Threshold and Weighted Fourier series.

Acknowledgements

I would like to thank my research advisor, Professor Moo K. Chung for his introduction to the field of statistics and medical imaging, and his guidance and encouragement during the entire course of my research. His passion about medical imaging, his rigorousness in mathematics, and his generousness in daily life helped me go through the most difficult time in my research and personal life. Without his support and help, I would have not made it this far.

I would like to thank Professor Andy Alexander, Professor Charles Dyer, Professor Vikas Singh, Professor Kam-Wah Tsui and Professor Grace Wahba, who serve as members of my Ph.D committee, for their helpful comments and suggestions. I would like to thank Professor Richard J. Davidson and Professor Kim M. Dalton for supporting the study of autism. I would like to thank Dr. Houri K. Vorperian for her supportive role throughout my graduate study.

I would like to thank my friends and my fellow students in the Department of Statistics, Weiliang Shi, Deyuan Jiang, Xiaolei Li, Xiaodan Wei, Huaibao Feng and Zhengxiao Wu. They made my life at Madison a wonderful journey. I also would like to thank my friend Jia Cao at Columbia University, and my colleague Christopher Tong at Merck for their illuminating discussions and suggestions. Their generous help of proofreading is crucial for the completion of my dissertation.

Finally I would like to thank my parents Guihe Wang and Meiying Sun, and my sisters Shuli Wang and Shuqin Wang, for their understanding and support for many years during all the twists and turns in my life. I also would like to thank my lovely nephews Hao Wen and Zheng Wang, who always bring smiles to my face even during a gloomy day. This dissertation is dedicated to them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Medical image analysis and acquisition techniques are experiencing an explosive growth due to the advancement of computer technology. Modern medical images provide physicians a remarkably detailed vision of the anatomical structures *in vivo*. This brings a dramatic increase of using medical images to help answer key questions that arise in human anatomical studies, disease diagnoses and drug development processes (Pien et al., 2005). For instance, various hard and soft tissue structures in the vocal tract area, whose measurements were unavailable in the past, can be measured at different ages using *magnetic resonance images* (MRI) and their growth patterns can be examined (Vorperian et al., 1999, 2005). Nacewicz et al. (2006) evaluated amygdala volumes using MRI and examined whether the variations in amygdala volume are related to the severity of autism. MRI has been used to study heart structure and function and to assess plaque composition and its regression in the coronary vasculature (Choudhury et al., 2002). Biochemical imaging biomarkers are being developed for the identification of "vulnerable plaque" for studies of primary prevention (Frank and Hargreaves, 2003). In these studies, the imaging biomarkers were treated as multivariate random variables. Multivariate statistical analysis can

be conventionally applied.

In the meantime, infinite-dimensional data, such as curves, surface and volumes, are also increasingly collected in medical image analysis. We usually refer to these infinite-dimensional data as functional data (Ramsay and Silverman, 1997, 2002). In practice, it is necessary for functional data analysis to achieve some form of dimension reduction so that one can reduce the infinite-dimensional data to finite and tractable dimensions. Fourier series decompose an L^2 function into a set of simple functions, which may be sines and cosines and complex exponentials. A cutting-off of high frequencies of a Fourier series usually gives a good smooth approximation of a periodic curve. Therefore, it has been applied to functional data analysis for the purpose of curve modeling and dimension reduction (Bracewell, 1999; Bosi and Goldberg, 2003). Recently, the *spherical harmonics* (SPHARM) (Sternberg and Smith, 1946; Hobson, 1955; Byerly, 1959), which is a higher dimensional Fourier series, has been widely applied to computer graphics and medical imaging for surface structure representations. Detailed remarks and historical references of SPHARM can be found in Groemer (1996). Brechbuehler et al. (1995) extended the concept of elliptical Fourier descriptor of closed curves and used a global parametrization to expand the object surface into a series of SPHARM functions. Gerig et al. (2001) and Shen et al. (2004) used SPHARM to represent the hippocampus and amygdala surfaces and statistical inference was made based on SPHARM representations. Kelemen et al. (1999), Gu et al. (2004) and Chung et al. (2006b) applied SPHARM

to characterize more complex cortical surfaces. Kazhdan et al. (2003) presented a novel tool that transforms rotation dependent shape descriptors into rotation independent SPHARM representations. The coefficients of SPHARM give a unique representation of the given anatomical structures. A direct application of this property can be found in Shen et al. (2004). They registered hippocampus surfaces based on the degree one SPHARM (an ellipsoid) and then applied *principal component analysis* (PCA) of the coefficients for detecting schizophrenia. In this dissertation, we refer to both classic Fourier series and SPHARM as Fourier series in general.

Even though Fourier Series have been widely used in medical image analysis, they have several drawbacks. Firstly, Gibbs (ringing) phenomenon occurs when using Fourier series to approximate a curve with sharp corners. Its oscillation patterns will not die with the increasing order of Fourier series as shown in Figure 1. The second drawback of Fourier series is that it is theoretically complicated and computationally time-consuming to estimate the smoothness of the approximation, which is crucial for the statistical inference using random field theory (Worsley, 1996; Cao and Worsley, 1999; Kiebel et al., 1999). One also has to be extremely careful about the normality assumption of the Fourier series-based generalized linear models. Improper choices of the degrees of Fourier series can cause violation of the normality assumption of estimated errors in the literature (Shen et al., 2004; Chung et al., 2008a).

We propose a novel systematic framework of *weighted Fourier series* (WFS)

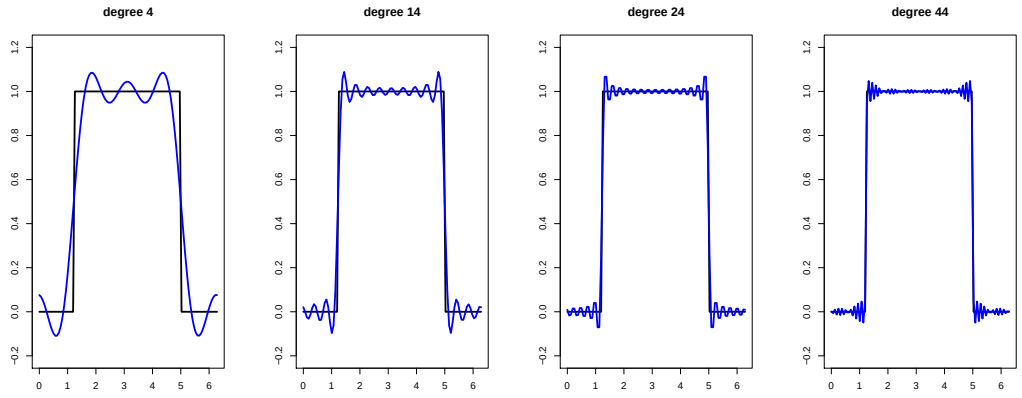

Figure 1: A demonstration of Gibbs phenomenon of Fourier expansions of degree 4, 14, 24, 44. The black curves are the original curve with sharp corners and the blue curves are the Fourier expansions of the original curve.

analysis (Chung, 2006a; Chung et al., 2006b, 2008a) that addresses many shortcomings associated with the traditional Fourier series analysis. WFS is closely related to heat kernel smoothing (Chung et al., 2005), which was applied as a novel data smoothing and analysis framework for cortical thickness data defined on the brain cortical manifold (Chung et al., 2005). It was pointed out that it is more natural to assign the weights based on the geodesic distance along the surface. A framework of using heat kernel smoothing detecting the regions of abnormal autistic cortical was developed via random field based multiple comparison correction. This paper built the ground work for the procedure of medical image analysis using kernel methods. WFS was first proposed and applied to the problem of detecting abnormal cortical regions in a clinical population by Chung et al. (2006a). For the smooth parametrization, they developed

a novel weighted spherical harmonic (SPHARM) representation. A theoretical framework for the weighted Fourier analysis was presented and how it could be used in the tensor-based morphometry was introduced. Chung and his colleagues also presented a novel multi-scale voxel-based morphometry using the WFS representation to address the optimal amount of registration that should be used in voxel-based morphometry (Chung et al., 2006b). Chung et al. (2007a) applied weighted Fourier analysis in quantifying the amount of gray matter in a group of high functioning autistic subjects. Most recently, Weighted Fourier series were also applied to detect abnormal cortical regions in the group of high functioning autistic subjects (Chung et al., 2008a). The authors also showed that a WFS is the least squares approximation to the solution of an isotropic heat diffusion on the unit sphere.

Even though weighted Fourier analysis of medical images is well defined theoretically and numerically, the implementation of Fourier series is not trivial as it looks, especially for the models that involve large data (e.g. cortical surfaces). Traditionally, Fourier series or SPHARM was derived from the least-squares estimation (LSE) (Gerig et al., 2001; Shen et al., 2004). But LSE requires the inversion of large matrices. The computation of large inverse matrices is in general very time-consuming. To deal with this problem, Shen and Chung (2006) and Chung et al. (2006b) proposed an *iterative residual fitting* (IRF) algorithm to improve the computational efficiency by decomposing the Hilbert space $L^2(S^2)$ into a direct product of a set of subspaces (i.e., by partitioning a

large design matrix into small submatrices in the linear model settings), then iteratively performing LSE using each small submatrix on the residuals. IRF greatly improves the computation efficiency. But IRF assumes that the submatrices are linearly independent pairwisely. In practice, this linear independency can not be achieved. The linear dependency between submatrices is always not negligible for the estimation of WFS. Therefore the tradeoff of fast computation of IRF is the loss of accuracy of the estimation. In this dissertation, we propose the *adaptive iterative regression* (AIR) method to address this issue. AIR inherits the idea of IRF by partitioning the function space into a set of subspaces. But AIR carries out an extra correction step to improve the orthogonality between two contiguous subspaces. The improved orthogonality reduces the bias in the WFS estimation. Our simulations show that computational efficiency of AIR is comparable with IRF, but its estimation is more accurate than IRF.

The studies of autism (Berument et al., 1999; Scott et al., 2002; Yeargin-Allsopp et al., 2003; Dalton et al., 2005a), recently attracted great interest in medical imaging studies. Autism is a neuro-developmental disorder affecting behavioral and social cognition, which manifests in delays of social interaction, language as used in social communication, or symbolic or imaginative play with onset prior to age 3 years. About 14 out of 10,000 children have autism or a related condition in the United States. The causes of autism are full of debates and controversy and there is no definite cure of autism. However, recent imaging studies showed connections between autism and various regions or tissue

structures of the brain, such as prefrontal cortex, medial and ventral temporal lobe, superior temporal sulcus, corpus callosum, amygdala hippocampus, cerebellum and so forth. Abell et al. (1999) used the voxel-based morphometry in high functioning autism to show decreased gray matter volume in the right paracingulate sulcus, the left occipito-temporal cortex, increased amygdala and periamygdaloid cortex. Vidal et al. (2003) showed reduced callosal thickness in the genu, midbody, and splenium in autistic children. Hoffmann et al. (2004) showed curvature difference in the midbody between autistic and normal subjects. Chung et al. (2004) applied a 2D version of voxel-based morphometry in differentiating the white matter concentration of the corpus callosum for the group of 16 high functioning autistic and 12 normal subjects. Dalton et al. (2005a) found that the activation in the fusiform gyrus and amygdala was strong and positively correlated with the time spent fixating the eyes in the autistic group. In Alexander et al. (2007), diffusion tensor measurements in corpus callosum were investigated in a large group of high-functioning autistic patients compared to matched controls.

To show the framework of WFS analysis of medical images, we are going to apply WFS analysis to the study of autism. Two data sets are used for the study:

- *Corpus Callosum data:* MR midsagittal slice images (as shown in Figure 2) of 15 high functioning autistic subjects and 12 normal controls. All subjects are right-handed males as shown in Figure 2.

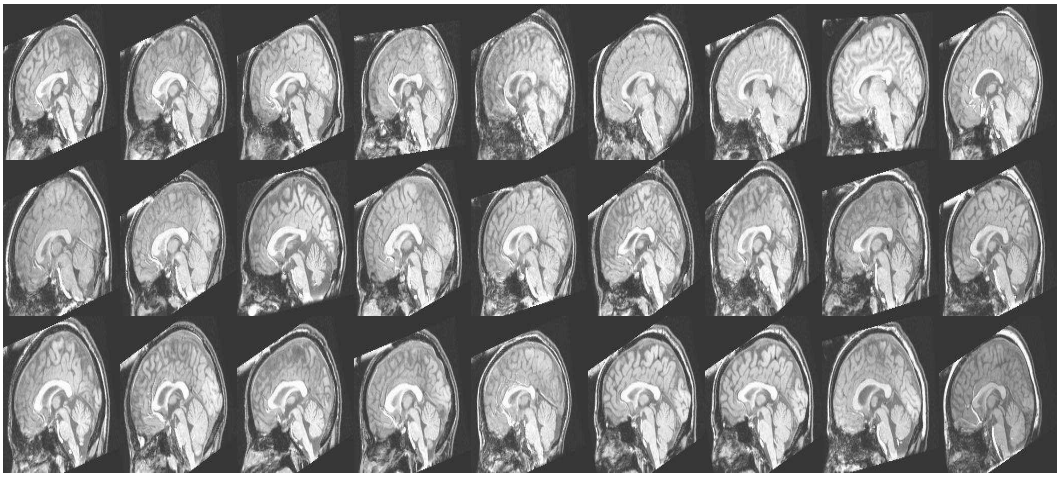

Figure 2: The corpus callosum data: all 27 mid-sagittal slice images, which include 15 high functioning autistic subjects and 12 normal controls.

- *Amygdala data:* MR volume images of 16 autistic subjects and 14 normal controls. Each subject includes a left and a right amygdala. There are total 60 images.

The two data sets were provided by the scientists from the Waisman Laboratory for Brain Imaging and Behavior at the University of Wisconsin. They were originally used to study the underlying relationships between autism and neuro-anatomical structures (Nacewicz et al., 2006). Cortical surface data and mandible surface data, which were provided by Waisman Laboratory for Brain Imaging and Behavior and Vocal Tract Development Laboratory at the University of Wisconsin, are also used for the illustration and simulations of our methods.

The general pipeline of WFS analysis framework is shown in Figure 3, which

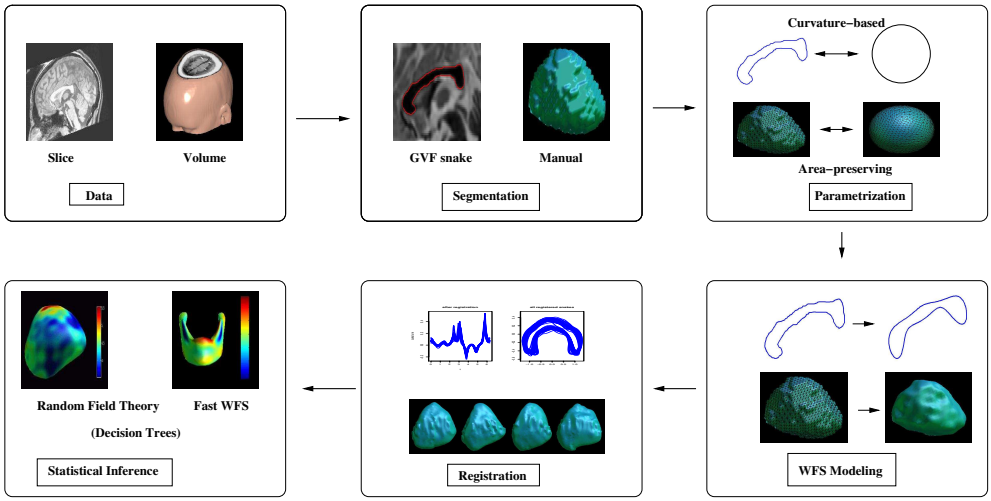

Figure 3: The pipeline of WFS analysis of medical images.

usually has the following steps: first, the boundaries of interest are extracted by manual or automatic segmentation methods; second, a parametrization procedure is proposed to find the optimal one-to-one mapping between the boundaries and unit sphere (parametrization) for mathematical modeling; WFS representations are calculated based on the parametrization results; a curvature-based affine alignment is applied, and then a curvature-based non-linear registration is carried out to further improve the registration results; final statistical analysis is made by using various tools and models.

The main contributions of this dissertation are:

- We extended the systematic theoretical framework of weighted Fourier analysis. We formulated the weighted Fourier analysis into the frame of the classic functional analysis and partial differential equations.

- We proposed an AIR method for the estimation of weighted Fourier series. This method was proved to be computationally efficient and numerically accurate by various simulations and studies.
- We proposed an AIR-based method to choose the optimal degrees of WFS using the F-statistics.
- We proposed a novel curve curvature estimation method, which is more robust and accurate than the finite difference method.
- We designed a curvature-based non-linear curve registration method.
- We proposed a WFS-based method of the surface curvature estimation and a curvature-based surface alignment method.
- We proposed a fast weighted Fourier analysis method, which provides fast estimation of WFS and chooses the significant frequencies automatically.

The structure of this dissertation is designed as follows: we briefly introduce the background of WFS, the basic content and structure of the dissertation in Chapter 1; in Chapter 2, we introduce the WFS representation as a solution to a Cauchy problem, and show that the WFS representation is a natural smoothing procedure, which not only improves the signal to noise ratio but also improves normality of the estimated errors; a series of important theoretical properties of WFS are stated and proved; we then numerically implement WFS with the AIR algorithm to improve the computational efficiency and accuracy; in Chapter 3, we design curvature-based curve and surface registrations based on the

proposed curvature estimation methods; in Chapter 4, we propose a novel fast weighted Fourier analysis method for WFS model selections; in Chapter 5, we apply the WFS image analysis techniques to the study of autism; we propose a decision tree-based automated diagnosis of autism using corpus callosum data; we find local difference between autistic and normal subjects in right amygdala by using random field theory; and we also apply fast weighted Fourier analysis to the study of growth patterns of mandible surfaces; Finally, we summarize our works in weighted Fourier analysis and discuss the possible approaches of future research in statistical shape analysis of anatomical structures in Chapter 6.

Chapter 2

Weighted Fourier Analysis

With technological advances in measurement devices and computational method, infinite-dimensional data, such as curves, surface and volumes, are increasingly collected in medical image analysis. We usually refer to these infinite-dimensional data as functional data (Ramsay and Silverman, 1997, 2002). Functional data analysis has to deal with functions and function spaces. Therefore, the concept of infinite-dimensional Hilbert space, L^2 in most cases, arises naturally and frequently for medical image analysis. Since this concept is a generalization of Euclidean space, geometric intuition plays an important role in many aspects of the Hilbert space theories. Analogous to Cartesian coordinates, an element of a Hilbert space can be uniquely characterized by its coordinates with respect to an orthonormal basis. In Euclidean spaces, the eigenvectors of a Hermitian matrix can be used to form such an orthonormal basis. Similar to the extension of vectors to functions, we replace matrices by linear operators in functional data analysis, in particular, Hermitian matrices are replaced by self-adjoint linear operators. In this chapter, we study a weighted Fourier series representation of the element in the Hilbert space based on the eigenfunctions of a self-adjoint linear operator. This weighted Fourier series representation can

be derived as a solution to the associated Cauchy problem.

2.1 Introduction to weighted Fourier series

2.1.1 The derivation of weighted Fourier series

In medical image analysis, one always deals with subjects that have a one-to-one mapping (isomorphism) to a circle, a sphere or a solid ball, which we consider as 1-dimensional, 2-dimensional or 3-dimensional unit spheres in the following context. Based on the one-to-one mapping, one considers the coordinates of these subjects as functions on the unit sphere, which encourages us to explore the characteristics of these functions and the properties of their related Hilbert spaces.

We start with a Hilbert space defined on a manifold (Stoker, 1969; Jost, 2002; Dragomir, 2006). A manifold is an abstract topological space in which every point locally resembles Euclidean space. Let $\mathcal{M} \in \mathbb{R}^d$ be a compact manifold. The squared-integrable function space, $L^2(\mathcal{M})$, is the Hilbert space defined on $\mathcal{M}$ with the inner product,

$$\langle f_1, f_2 \rangle = \int_{\mathcal{M}} f_1(x) f_2(x) d\mu(x), \quad \text{for any } f_1, f_2 \in L^2(\mathcal{M}),$$

where μ is the Lebesgue measure defined on $\mathcal{M}$. The proof of the completeness of $L^2(\mathcal{M})$ is a classic result in functional analysis (Halmos, 1978; Conway, 1985; Rubin, 1991). In addition, $L^2(\mathcal{M})$ is separable. Therefore, any element in

$L^2(\mathcal{M})$ can be represented by a countable number of elements. This property guarantees the existence of the countable orthonormal basis.

For seeking an appropriate orthonormal basis, non-degenerate self-adjoint linear operators on $L^2(\mathcal{M})$ are of special interest. On a finite-dimensional inner product space, a self-adjoint operator $\mathcal{L}$ can be defined by its corresponding Hermitian matrix $M_{\mathcal{L}}$ ($M_{\mathcal{L}}$ is equal to its conjugate transpose). By similarity transformation,

$$M_{\mathcal{L}} = U^{-1}\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)U \tag{1}$$

where U is the unitary matrix whose columns are the eigenvectors of $M_{\mathcal{L}}$ and $\lambda_j, j = 1, 2, \cdots, n$ are the eigenvalues of $M_{\mathcal{L}}$. The operator $\mathcal{L}$ (or matrix) can be represented as a diagonal matrix $\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ with entries in the real numbers in the space spanned by the columns of U. The self-adjoint operators on infinite dimensional Hilbert spaces essentially resemble their finite dimensional counterparts.

A linear operator $\mathcal{L} : L^2(\mathcal{M}) \to L^2(\mathcal{M})$ is *self-adjoint* if

$$\langle \mathcal{L}f_1, f_2 \rangle = \overline{\langle f_1, \mathcal{L}f_2 \rangle},$$

where the overline indicates the complex conjugate. From the definition, self-adjoint operators are "symmetric". Just like symmetric matrices, self-adjoint operators can be diagonalized. Therefore a self-adjoint operator can be determined completely by its eigenvalues and eigenfunctions. In particular, these eigenvalues are real. Let $\{\lambda_i\}$ and $\{\phi_i\}$ $(i = 1, 2, \cdots)$ be the eigenvalues and

eigenfunctions of $\mathcal{L}$ such that

$$\mathcal{L}\phi_i = \lambda_i \phi_i.$$

Then $\{\phi_i\}_{i=1}^{\infty}$ is a complete orthonormal basis of $L^2(\mathcal{M})$. Similar to (1), one can write a self-adjoint operator in the form of a Hilbert-Schmidt kernel (Courant and Hilbert, 1953; Berezankii, 1968),

$$K_{\mathcal{L}}(p,q) = \sum_{i=1}^{\infty} \lambda_i \phi_i(p)\phi_i(q).$$

This is the infinite-dimensional version of $M_{\mathcal{L}}$ in (1).

For Hilbert spaces, one common choice of basis is the Fourier basis. Under moderate computation, one can derive the Fourier series and *spherical harmonics* (SPHARM) as the eigenfunctions of a self-adjoint operator, the negative of the Laplacian $\mathcal{L} = -\triangle$ (which makes the operator non-negative), defined on the unit sphere. If $\mathcal{M} = S^1$, the unit circle, then

$$\mathcal{L} = -\partial^2/\partial\theta^2,$$

where $\mathcal{L}$ has eigenvalues $\{l^2, l = 0, 1, \cdots, \}$. That $\mathcal{L}f_{li} = l^2 f_{li}$, $i = 1, 2$ derives the Fourier basis

$$f_0 = \frac{1}{\sqrt{2\pi}}, \quad f_{l1} = \frac{\sin l\theta}{\sqrt{\pi}}, \quad f_{l2} = \frac{\cos l\theta}{\sqrt{\pi}}, \quad l = 1, 2, \cdots,$$

where $\theta \in [0, 2\pi]$. Similarly, if $\mathcal{M} = S^2$, SPHARM can be derived as a solution to the system

$$\begin{cases} \triangle Y_{lm} &= \lambda_l Y_{lm}, \quad l = 1, 2, \cdots, \quad -l \le m \le l, \\ \triangle &= \frac{\partial}{\sin\theta \partial\theta}(\sin\theta \frac{\partial}{\partial\theta}) + \frac{\partial^2}{\sin^2\theta \partial^2\phi}, \\ \lambda_l &= l(l+1), \end{cases}$$

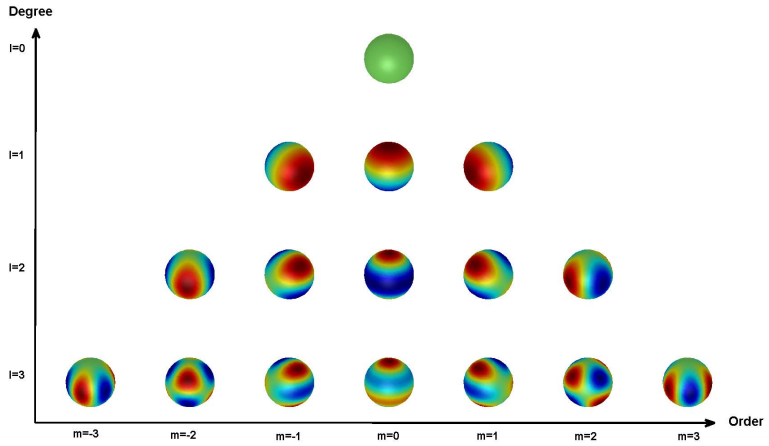

Figure 4: Plots of SPHARM basis functions of degrees from 0 to 3. The color indicates the magnitude of the function. The x-axis and y-axis show the correspondence of the degrees and the orders of the SPHARM basis functions.

which is

$$Y_{lm} = \begin{cases} \sqrt{\frac{(2l+1)(l-|m|)!}{2\pi(l+|m|)!}} P_l^{|m|}(\cos\theta)\sin(|m|\phi), & -l \le m \le -1, \\ \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}} P_l^0(\cos\theta), & m = 0 \\ \sqrt{\frac{(2l+1)(l-|m|)!}{2\pi(l+|m|)!}} P_l^{|m|}(\cos\theta)\cos(|m|\phi), & 1 \le m \le l, \end{cases}$$

where $\theta \in [0, \pi]$ is the zenith angle (also known as polar angle), which starts from the z-axis, and $\phi \in [0, 2\pi]$ is the azimuth angle, which starts from the x-axis, and $P_l^{|m|}$ is the associated Legendre functions of degree l and order m. Y_{lm} is called the SPHARM of degree l and order m. SPHARM basis functions of degree 0 to 3 are plotted in Figure 4, which shows that the distribution of SPHARM is in the form of a pyramid. Gerig et al. (2001); Bulow (2004); Gu et al. (2004); Shen et al. (2004) used the complex-valued SPHARM, which is

from the original definition of spherical harmonics. Even though real-valued and complex-valued SPHARMs are essentially equivalent, the coefficients of the real-valued SPHARM are more meaningful and interpretable for the generalized linear models that we will specify later.

Fourier series was invented to express the solution of the heat equation (Fourier, 1822). In our work, we are going to introduce weighted Fourier series as a solution to the Cauchy problem, a generalized form of the heat equation. Suppose we have a smooth manifold $\mathcal{M}$ ($\mathcal{M}$ is called a Cauchy surface). A Cauchy problem consists of finding the solution $g(p,t)$ of the differential equation which satisfies

$$\begin{cases} \frac{\partial g(p,t)}{\partial t} + \mathcal{L}g(p,t) &= 0, \qquad t \geq 0, \quad p \in \mathcal{M} \\ g(p,0) &= f(p). \end{cases} \tag{2}$$

Equation (2) becomes a heat equation with given initial condition when $\mathcal{L} = -\triangle$. Equation (2) defines a natural smoothing procedure with input function $f(p)$ (the initial condition). t controls the amount of smoothing and is termed as the *bandwidth*. The existence and uniqueness of the solution to the Cauchy problem is stated in the following theorem, which was first presented and proven in Chung et al. (2007a).

Theorem 2.1. *Given that the eigenvalues* $\{\lambda_j\}_{j=1}^{\infty}$ *and eigenfunctions* $\{\phi_j\}_{j=1}^{\infty}$ *of* $\mathcal{L}$ *are known, the unique solution to (1) is given as*

$$g(p,t) = \sum_{j=0}^{\infty} e^{-\lambda_j t} \langle f, \phi_j \rangle \phi_j(p). \tag{3}$$

if $\mathcal{L}$ *is non-degenerate, compact and self-adjoint.*

Proof. The Cauchy-Kowalevski theorem (Cauchy, 1842; Kowalevski, 1875; Gorbachuk, 1998; Nakhushev, 2001) gives the proof of the uniqueness and existence of the Cauchy problem for a general linear operator, $\mathcal{L}$. In this proof, only self-adjoint operators are considered. If a self-adjoint linear operator, $\mathcal{L}$, has non-zero eigenvalues, then it is non-degenerate, i.e., it has infinitely many eigenfunctions and its eigenfunctions consist of a complete basis of $L^2(\mathcal{M})$ (Aupetit, 1991). Since $\{\phi_j\}_{j=1}^{\infty}$ are complete and orthonormal,

$$g(p,t) = \sum_{j=0}^{\infty} \langle g(x,t), \phi_j \rangle \phi_j(p).$$

By $\mathcal{L}\phi_j = \lambda_j \phi_j$, then equation (2) becomes

$$\partial_t (\sum_{j=0}^{\infty} \langle g, \phi_j \rangle \phi_j(p)) = \sum_{j=0}^{\infty} \lambda_j \langle g, \phi_j \rangle \phi_j(p)$$

where the exchangeability of differentiation and summation is based on the fact that a Fourier series is uniformly convergent in L^2 (Rudin, 1976). By the orthonormality of $\{\phi_j\}_{j=1}^{\infty}$, we have

$$\partial_t (\langle g, \phi_j \rangle \phi_j(p)) = \lambda_j \langle g, \phi_j \rangle \phi_j(p), \quad j = 1, 2, \cdots$$

Now one only needs to solve a much simpler partial differential equation for each j that has the form as $\partial_t g + \lambda g = 0$ with initial condition $g(x,0) = f$. The solution simply is $g(x,t) = e^{-\lambda t} f$. Therefore,

$$\langle g(x,t), \phi_j \rangle \phi_j(p) \quad = \quad e^{-\lambda_j t} \langle f, \phi_j \rangle \phi_i(p).$$

By putting all terms together, we have

$$g(p,t) \quad = \quad \sum_{j=0}^{\infty} e^{-\lambda_j t} \langle f, \phi_j \rangle \phi_i(p),$$

which is a solution to (2).

To prove the uniqueness of the solution, let $\tilde{f} = \sum_{j=0}^{\infty} a_j \phi_j(p)$. Then, we plug $\tilde{f} - g(p,t)$ into (2) to get $a_j = e^{-\lambda_j t}\langle f, \phi_j \rangle$ for every j, which shows that the solution is unique. □

We call $g(p,t)$ the *weighted Fourier series* (WFS) of function f since it has an extra weight term for every coefficient comparing with the Fourier series representation. Similar to heat kernel smoothing (Chung, 2006b), WFS provides a method of kernel smoothing. It is easy to verify that WFS has the basic properties of a smoothing process.

Theorem 2.2. *Let $\{\phi_i\}_{i=0}^{\infty}$ be a Fourier basis or SPHARM basis and assume f is bounded on the compact support $\mathcal{M}$. If the bandwidth $t \to 0$, WFS defined in (3) converges to a Fourier series or SPHARM representation pointwisely*

$$\lim_{t\to 0} g(p,t) \;=\; \sum_{i=0}^{n} \langle f, \phi \rangle \phi_i, \quad \textit{for every } p,$$

and

$$\lim_{t\to \infty} g(p,t) \to \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} f(p) d\mu(p), \quad \textit{for every } p.$$

Proof. We first prove that $g(p,t)$ defined in (3), pointwisely converges to its Fourier series representation as $t \to 0$. Since $\|\phi_i\|_2 = 1$ and $\mathcal{M}$ is compact, ϕ_i is bounded on $\mathcal{M}$. Note that as $\lambda_i \to \infty$ as $i \to \infty$. By Hölder's inequality,

$$\begin{aligned} |e^{-\lambda_i t}\langle f, \phi \rangle \phi_i| \;&\leq\; |e^{-\lambda_i t}\phi_i| \cdot \|f\|_2 \cdot \|\phi_i\|_2 \\ &=\; C_0 e^{-\lambda_i t}, \end{aligned}$$

where C_0 is a constant that is independent of i. Since $\sum_{i=0}^{\infty} C_0 e^{-\lambda_i t}$ is convergent. Then by bounded convergence theorem (Rudin, 1976), for any fixed p, one can switch the limit and the summation

$$\begin{aligned}\lim_{t\to 0} g(p,t) &= \sum_{i=0}^{\infty} \lim_{t\to 0} e^{-\lambda_i t}\langle f,\phi\rangle\phi_i \\ &= \sum_{i=0}^{\infty}\langle f,\phi\rangle\phi_i,\end{aligned}$$

and

$$\begin{aligned}\lim_{t\to \infty} g(p,t) &= \sum_{i=0}^{\infty} \lim_{t\to \infty} e^{-\lambda_i t}\langle f,\phi\rangle\phi_i \\ &= \langle f,\phi_0\rangle\phi_0 \\ &= \frac{1}{\mu(\mathcal{M})}\int_{\mathcal{M}} f d\mu.\end{aligned}$$

□

This theorem also tells us that the Fourier series is a special case of WFS (with bandwidth 0). Thus with an appropriately chosen bandwidth, WFS is usually a better choice than Fourier series.

2.1.2 The heat kernel

WFS is directly related to heat kernel smoothing. The heat kernel is the generalization of the Gaussian kernel defined in the Euclidean space to an arbitrary Riemannian manifold (Rosenberg, 1997; Chung et al., 2005, 2007a). We can construct the heat kernel on the compact Riemannian manifolds and represent

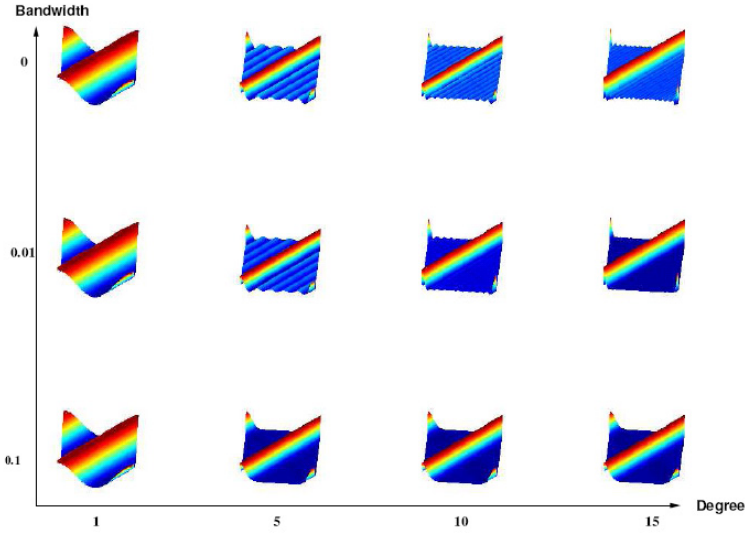

Figure 5: Plots of heat kernel $K_t^k(p,q)$ on S^1 with degree = 1, 5, 10, 15 for every bandwidth t=0, 0.01, 0.1, where $(p,q) \in [0,2\pi] \times [0,2\pi]$.

the heat kernel as

$$K_t(p,q) = \sum_{i=1}^{\infty} e^{-\lambda_i t}\phi_i(p)\phi_i(q). \quad (4)$$

In practice, heat kernels with finite terms,

$$K_t^k(p,q) = \sum_{i=1}^{k} e^{-\lambda_i t}\phi_i(p)\phi_i(q),$$

are often used to approximate the underlying heat kernel. Here k is called the degree of the heat kernel. Figure 5 shows plots of heat kernels on S^1 for different degrees with different bandwidths. From this figure, one can also see that WFS gives a good smooth approximation of the heat kernel with different bandwidths. Selecting the optimal degree and bandwidth of a WFS kernel will be an interesting topic. Generalized cross-validation (GCV) (Wahba, 1990) and

the discrepancy principle (DP) (De Nicolao et al., 1997; Sparacino et al., 2001; Toffolo et al., 2001) can be good candidates for certain cases. But GCV and DP are in general computationally expensive for large image data. We are going to address this issue using an F-statistics based model selection method.

WFS kernel is indeed an integral kernel. One can define a heat kernel smoothing operator $T : L^2(\mathcal{M}) \to L^2(\mathcal{M})$ as

$$T_t(f(p)) = \int_{\mathcal{M}} K_t(p, q) f(q) d\mu(q).$$

By Ascoli-Arzela theorem (Rubin, 1991), one can prove that the operator T is compact and self-adjoint. Therefore, the heat equation becomes a special case of the famous Sturm-Liouville problem with initial conditions. The WFS representation of initial condition f is automatically a solution to (2) as

$$g(p, t) \quad = T_t(f(p)) \quad = \int_{\mathcal{M}} K_t(p, q) f(q) d\mu(q).$$

For any fixed q, $K_t(p, q)$ is a probability distribution function centered at q, which is also shown in Figure 5. One can also easily check that

$$\int_{\mathcal{M}} K_t(p, q) du(q) \quad = T_t(1) \quad = 1$$

where the second equality is derived from the fact that the WFS of 1 is 1. This coincides with Gaussian kernel smoothing.

Furthermore, using the harmonic addition theorem (Wahba, 1990; Chung et al., 2007b), one can further simplify the heat kernel on S^2 as

$$K_t(p, q) = \sum_{l=0}^{k} \frac{2l + 1}{4\pi} e^{-l(l+1)t} P_l^0(\cos \gamma) \tag{5}$$

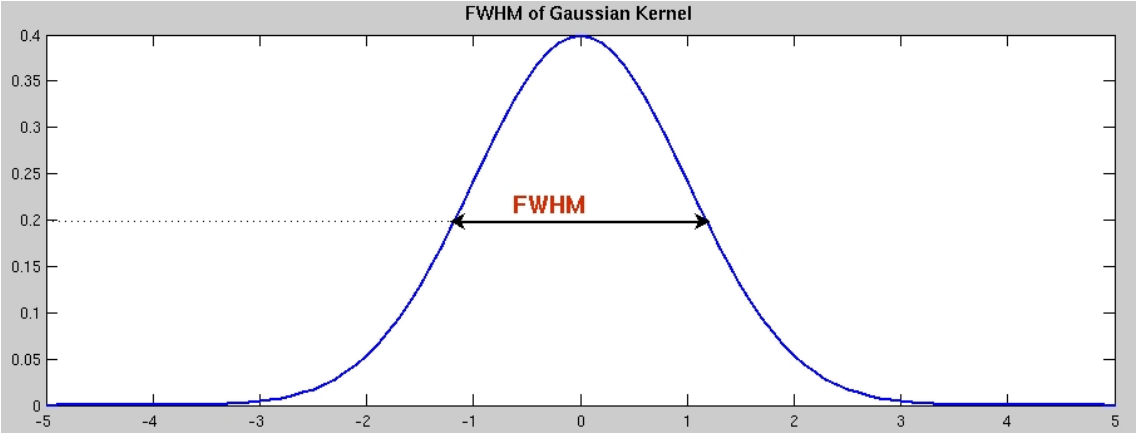

Figure 6: The FWHM of Gaussian kernel.

where γ is the angle between p and q. This step will make the calculation of the *full width at half maximum* (FWHM) of the heat kernel relatively easy. The FWHM is very important to characterize the smoothness of images in random field theory (Worsley, 1996; Cao and Worsley, 1999). The FWHM of a function is given by the difference between the two extreme values of the independent variable at which the dependent variable is equal to half of its maximum value. The FWHM of a Gaussian kernel (as shown in Figure 6) can be explicitly given as

$$\text{FWHM} = 2\sqrt{\log 2}\sigma.$$

where σ is the bandwidth of the Gaussian kernel.

To calculate the FWHM of the heat kernel, we fix p in equation (5) to be the north pole and vary $\gamma = \cos^{-1}(pq)$. The maximum is obtained at $\gamma = 0$.

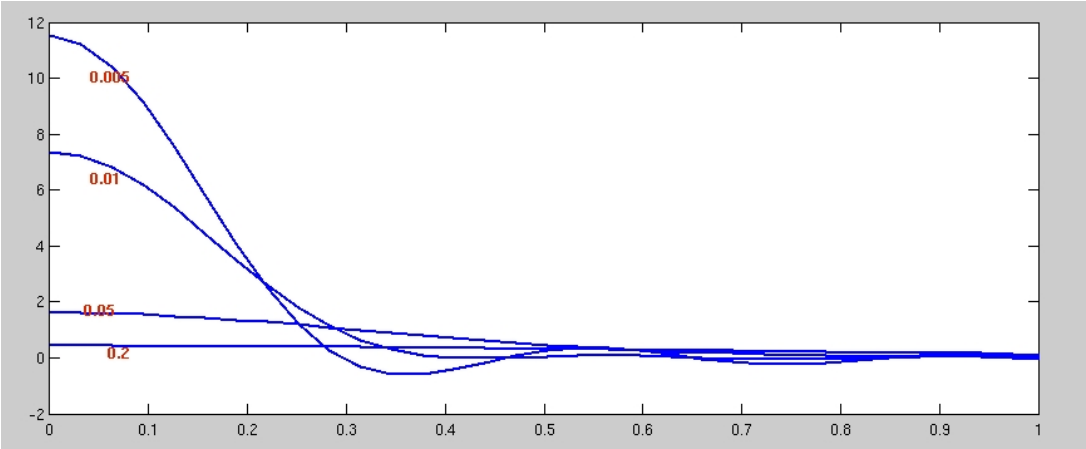

Figure 7: The heat kernels with t =0.005, 0.01, 0.05, 0.2, and k=15.

The FWHM is solved numerically for γ in

$$\frac{1}{2}\sum_{l=0}^{k} e^{-l(l+1)t} \cdot \frac{2l+1}{4\pi} = \sum_{l=0}^{k} e^{-l(l+1)t} \cdot \frac{2l+1}{4\pi} P_l^0(\cos\gamma). \tag{6}$$

The heat kernels with different bandwidths are shown in Figure 7. The relationship between FWHM and bandwidth t can be derived from equation (6). Similarly to Gaussian kernel, the larger the bandwidth of weighted Fourier kernel, the larger FWHM.

2.1.3 Reduction of Gibbs phenomenon

It is well-known that approximating a discontinuous function by Fourier series results in poor accuracy due to Gibbs phenomenon (a review, general definition and analysis of Gibbs phenomenon can be found in Gottlieb and Shu (1997)).

In Chapter 1, we pointed out that Gibbs (ringing) phenomenon happens

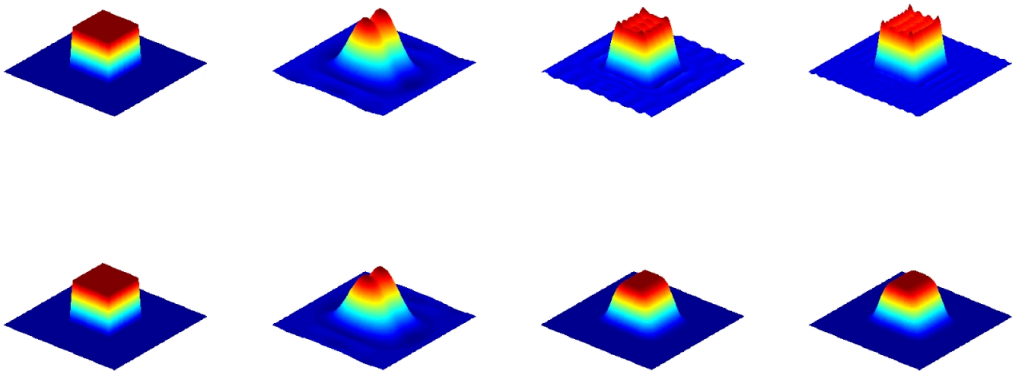

Figure 8: The plots demonstrate that WFS reduces Gibbs phenomenon. The first column shows the plots of a step function defined on $(\theta, \phi) \in [0, \pi] \times [0, 2\pi]$, where this function is 1 if $(\theta, \phi) \in [\frac{1}{3}\pi, \frac{2}{3}\pi] \times [\frac{2}{3}\pi, \frac{4}{3}\pi]$, and 0 elsewhere. The 2nd to 4th plots of the first row are SPHARM representations of the defined step function with degrees 5, 15, 25. The 2nd to 4th plots of the second row are the WFS representations of the defined step function with degrees 5, 15, 25 and bandwidth 0.01.

when using Fourier series to approximate a curve with sharp corners. Its oscillation patterns will not die with the increasing order of Fourier series as shown in Figure 1. The typical images in applications have sharp contours giving rise to discontinuities in the image functions. Gibbs phenomenon also happens when using SPHARM to approximate a surface that has sharp corners (as shown in the first row in Figure 8).

The following lemma (Gottlieb and Shu, 1997; Bronstein et al., 2002) mathematically characterizes the Gibbs phenomenon.

Lemma 2.1. *Assume that we have a piecewise continuous function* $f(x)$, $x \in$

$[0, 2\pi]$. *Let* $\{(a_k, b_k)\}_{j=0}^{K}$ *be the Fourier coefficients of* $f(x)$. *Then we have*

$$\max_{x\in[0,2\pi]} |f(x) - \sum_{j=0}^{K}\{a_k \cos(kx) + b_k \sin(kx)\}| \quad = \quad \frac{DP}{2\pi},$$

where $D = \max_x |f(x+) - f(x-)|$ *and*

$$P \quad = \quad \int_0^{2\pi} \frac{sinx}{x} dx.$$

Methods have been proposed to reduce the Gibbs phenomenon. Gottlieb et al. (2000) proposed a new filter in Fourier space to enhance the accuracy away from the discontinuities. Bronstein et al. (2002) proposed medical image reconstruction algorithm that makes use of forward nonuniform fast Fourier transform (NUFFT) for iterative Fourier inversion. Incorporation of total variation regularization allows the reduction of noise and Gibbs phenomena while preserving the edges.

Therefore, an efficient way to reduce Gibbs phenomenon is to use a smoothing procedure. As we can see from the definition, with the increasing degrees of WFS, the weights are getting smaller, which means WFS reduces the amount of high frequent noise. This property leads to one major advantage of WFS over Fourier series: WFS can effectively reduce Gibbs phenomenon. Note that WFS smoothing requires minimal amount of extra computation if the Fourier series representations are available.

To show this advantage of WFS, we define a step function on $(\theta, \phi) \in [0, \pi] \times [0, 2\pi]$. This function is 1 if $(\theta, \phi) \in [\frac{1}{3}\pi, \frac{2}{3}\pi] \times [\frac{2}{3}\pi, \frac{4}{3}\pi]$, 0 elsewhere, which is shown in first column of Figure 8. The degree 15 and 25 SPHARM

representations have spikes around the corners, while the corresponding WFS representations show no oscillated patterns and give a better smooth approximation of the pre-specified step function.

2.1.4 The normality of assumption

For the convenience of setting up Fourier series-based models and WFS-based models and performing hypothesis tests on medical images, the normality of errors is usually assumed (Shen and Chung, 2006; Chung et al., 2008a). To apply random field theory for image analysis, normality of errors is also assumed (Worsley, 1996; Cao and Worsley, 1999; Chung et al., 2008a).

Given an observation (a curve or a surface) $\boldsymbol{f}$, we want to represent it using Fourier series or WFS representations. In general, one pre-specifies a subspace $\mathcal{H}_K$ of $L^2(\mathcal{M})$ with proper dimension K (Shen et al., 2004; Chung et al., 2006a). We consider the following model,

$$E\boldsymbol{f}(p) = \sum_{i=1}^{K} e^{-\lambda_i t}\beta_i\phi_i(p),$$

where $p \in \mathcal{M}$. And the coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_K)$ are estimated from the linear model,

$$\boldsymbol{f} = \boldsymbol{Y}\boldsymbol{\Lambda}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2\boldsymbol{I}), \tag{7}$$

where $\boldsymbol{\Lambda} = \text{diag}(e^{-\lambda_1 t}, e^{-\lambda_2 t}, \cdots, e^{-\lambda_K t})$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_K)$ are the coefficients of the Fourier representation, and the design matrix of this linear model

is

$$\boldsymbol{Y} = \begin{bmatrix} \phi_1(p_1) & \cdots & \phi_K(p_1) \\ \vdots & \ddots & \vdots \\ \phi_1(p_n) & \cdots & \phi_K(p_n) \end{bmatrix}. \tag{8}$$

Here $\{\phi_i\}_{i=1}^K$ are discrete Fourier basis functions.

To show that WFS representations improve the normality assumption in Equation (7), we fit a linear model to a noisy amygdala surface from the autism study (Nacewicz et al., 2006), using SPHARM basis and apply the estimated coefficient $\boldsymbol{\beta}$ to both SPHARM and WFS representations based on an observed amygdala surface. We then plot the normal Quantile-Quantile (QQ) graphs of the estimated errors to assess the normality assumption for the fittings of different degrees. In Figure 9, we show that, for SPHARM representation, one always needs to find the proper degree (degree 15) to satisfy the normality assumption of the noise. Either over-smoothing (lower degrees) or over-fitting (higher degrees) will give a severe violation of the normality assumption, which is shown by skewed patterns in the QQ-plots. On the other hand, the normality assumption is still valid even if WFS representations have higher degrees, in which case the SPHARM representation will exhibit over-fitting.

In conclusion, WFS has the following properties and advantages (over Fourier series)

- WFS is both a fitting procedure and a smoothing procedure. Fourier series is a special case of WFS;

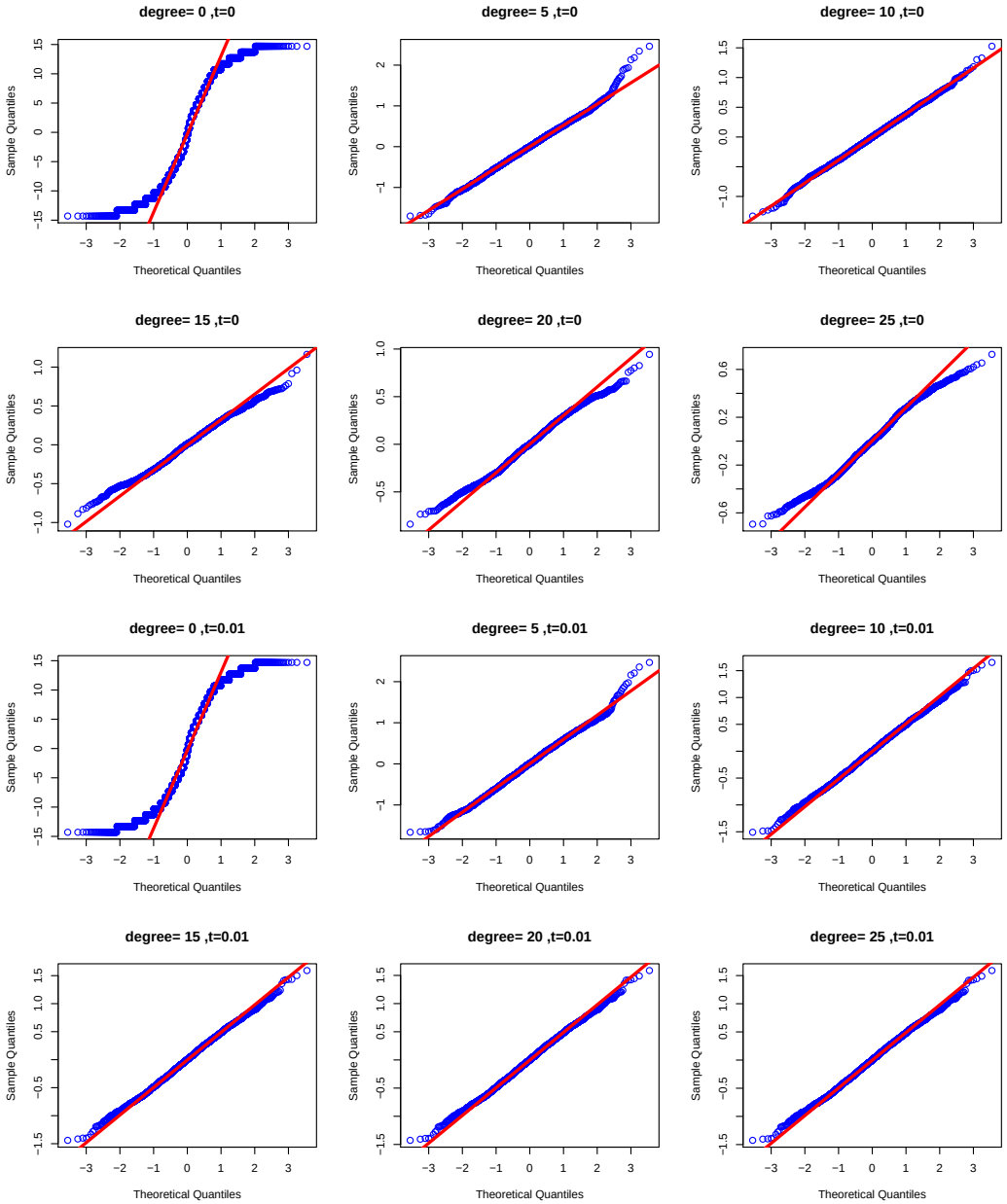

Figure 9: The plots for the test of normality and an amygdala surface from the study of autism is used for the demonstration. The first two rows are the quantile-quantile (QQ) plot of Fourier Series (SPHARM)-based linear models using degrees 0, 5, 10, 15, 20, 25. The last two rows are the QQ-plots of WFS-based linear models with bandwidth 0.01.

- WFS reduces the Gibbs phenomenon in Fourier series approximation;
- WFS is robust for the normality assumption in its related linear models;
- It is relatively easy to compute the smoothness of the WFS kernel in applying the random field theory (Worsley, 1996; Cao and Worsley, 1999).

2.2 Adaptive iterative regression

2.2.1 Least squares estimation and stepwise regression

To estimate the coefficients of WFS, one usually minimizes the *mean squared errors* (MSE),

$$\text{MSE}(\boldsymbol{\beta}) = (\boldsymbol{f} - \boldsymbol{Y}\boldsymbol{\beta})'(\boldsymbol{f} - \boldsymbol{Y}\boldsymbol{\beta}). \tag{9}$$

MSE can also be considered as the discrete L^2-distance, which gives this minimization a natural interpretation in functional analysis in the Hilbert space $L^2(\mathcal{M})$. The estimator that minimizes $\text{MSE}(\boldsymbol{\beta})$ in equation (9) is called the *least squared error* (LSE) estimator. By checking the following conditions for an optimization,

$$\begin{aligned} \tfrac{d}{d\boldsymbol{\beta}}\text{MSE}(\boldsymbol{\beta}) &= 0, \\ \tfrac{d^2}{d\boldsymbol{\beta}^2}\text{MSE}(\boldsymbol{\beta}) &> 0, \end{aligned} \tag{10}$$

or just simply checking the first equation in (10) by using the fact that MSE is positive and quadratic, the LSE of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Y}'\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{f}. \tag{11}$$

$\hat{\boldsymbol{\beta}}$ is also a maximum likelihood estimator (MLE) under the normality assumption.

An LSE is in general an optimal, unbiased and robust estimator (Bickel and Doksum, 2000; Shao, 2003) as shown in the following lemma.

Lemma 2.2. *Let $\hat{\boldsymbol{\beta}}$ be the LSE of (11).*

1. *If $\boldsymbol{\epsilon}$ are normally, independently and identically distributed (i.i.d.), $\hat{\boldsymbol{\beta}}$ is the uniformly minimum variance unbiased estimator (UMVUE).*

2. *If $\boldsymbol{\epsilon}$ are i.i.d., $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator (BLUE).*

3. *A BLUE is always robust.*

Besides all these good properties in Lemma 2.2, LSE is also numerically straightforward to implement. However, for medical image analysis, the observation $\boldsymbol{f}$ in (11) can be extremely large. For example, the number of vertices of a brain surface mesh can be larger than 40,000 (Shen and Chung, 2006; Chung et al., 2007b). The physical memory to store the large design matrices alone can easily reach the limits of most personal computers. It requires as many as 7,000 SPHARM basis functions (the columns of $\boldsymbol{Y}$ in (11)) to give a good representation of this cortical surface. The numerical operation of the design matrix with dimension as large as $40,000 \times 7,000$ can not be processed directly in the physical memory of a personal computer, which also makes it conceivably difficult to compute the inverse of the large matrix in (11). To overcome the computational difficulty, alternative methods have been developed.

Stepwise regression methods attracted a lot of attention more than 40 years ago (Freund et al., 1961; Goldberger, 1961; Goldberger and Jochemes, 1961). It can be potentially applied for solving large linear systems. For the stepwise regression, one first partitions the design matrix $\boldsymbol{Y}$ into two submatrices, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$. Rather than fitting the full model once and for all, one fits the simpler model,

$$\boldsymbol{f} = \boldsymbol{Y}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1.$$

In the second step, one fits the residual $\boldsymbol{\epsilon_1}$ using the second submatrix,

$$\boldsymbol{\epsilon}_1 = \boldsymbol{Y}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2.$$

Then the full model will be

$$\boldsymbol{f} = \boldsymbol{Y}_1\boldsymbol{\beta}_1 + \boldsymbol{Y}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2.$$

This two-step procedure was originally referred to as stepwise least squares (Goldberger, 1961) or residual analysis (Freund et al., 1961). The relationship between the estimation of $\boldsymbol{\beta}_2$ using a stepwise regression model and the full model was derived by Freund et al. (1961) and Goldberger and Jochemes (1961). They showed that stepwise regression always underestimates $\boldsymbol{\beta}_2$ in absolute value. By not realizing the increasing complexity and size of the data with the advancement of the high-speed computer, Alley (1987) falsely claimed that "Prior to the advent of the high-speed computer, stepwise regression was used at times as a simple method of estimating β's in multiple regression. Stepwise

regression is of limited value as a technique in today's world of high-speed computers". Not only is a stepwise regression needed for analysis of large medical image data (Shen and Chung, 2006; Chung et al., 2007b), but stepwise regression is also valuable for the selection of important predictors when the basis functions are redundant. For example, a recent algorithm, matching pursuit (Mallat and Zhang, 1993), decomposes any time-dependent signal to a linear expansion of waveforms that are selected from a redundant dictionary of functions by iteratively minimizing the residuals. Selecting the most important waveforms simultaneously is impossible since there are so many (in fact, uncountable) basis functions to choose from that the computation becomes infeasible.

But a two-step regression does not necessarily make the estimation of WFS coefficients simpler enough to carry out for large data such as cortical surfaces. Shen and Chung (2006); Chung et al. (2007b) generalized two-step regression to a K-stepwise regression fashion, which they called *iterative residual fitting* (IRF). The IRF procedure is described as following:

1. Partition the design matrix into submatrices as $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_K)$, where submatrix $\boldsymbol{Y}_i$ is a set of consecutive columns of $\boldsymbol{Y}$.

2. Regress $\boldsymbol{f}$ on the first submatrix $\boldsymbol{\beta}_1 = (\boldsymbol{Y}_1'\boldsymbol{Y}_1)^{-1}\boldsymbol{Y}_1'\boldsymbol{f}$. Save the first residual vector, $\boldsymbol{e}_1 = \boldsymbol{f} - \boldsymbol{Y}_1\boldsymbol{\beta}_1$.

3. For $1 \leq j < K$, compute the coefficients on the submatrix $\boldsymbol{Y}_{j+1}$,

$$\boldsymbol{\beta}_{j+1} \;=\; (\boldsymbol{Y}_{j+1}'\boldsymbol{Y}_{j+1})^{-1}\boldsymbol{Y}_{j+1}'\boldsymbol{e}_j$$

and calculate the j-th residual

$$\boldsymbol{e}_{j+1} \quad = \quad \boldsymbol{e}_j - \boldsymbol{Y}_{j+1}\boldsymbol{\beta}_{j+1}.$$

4. The estimation of the coefficients is

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \cdots, \boldsymbol{\beta}_K')',$$

and our fit will be

$$\hat{\boldsymbol{f}} = \sum_{j=1}^{K} \boldsymbol{Y}_j \boldsymbol{\beta}_j.$$

Simple calculations can show that IRF is computationally more efficient than LSE. For a design matrix $\boldsymbol{Y}$ with dimension $N \times P$, LSE needs to compute the inverse of $\boldsymbol{Y}'\boldsymbol{Y}$, whose dimension is $P \times P$. For the most widely used algorithms of matrix inversion, such as Gauss-Jordan elimination (Lipschutz and Lipson, 2001; Strang, 2003), LU decomposition (Horn and Johnson, 1985; Okunev and Johnson, 1997), QR decomposition (Becker et al., 1988) and so forth, the arithmetic computation is $O(P^3)$. For IRF, one needs to compute the inverse of K submatrices with dimension $P/K \times P/K$. Therefore, the arithmetic computation for IRF is $O(K \times (P/K)^3)$, i.e. $O(P^3/K^2)$. So for $K \geq 2$, the computation of IRF is always faster than that of LSE. For large K, the computational efficiency can be improved dramatically by IRF.

2.2.2 Adaptive iterative regression

As we are going to show in the later context, IRF is computationally efficient by being exempted from putting the entire design matrix into the physical memory

of the computer, and free of calculating the inversion of large matrices. But IRF estimation is always biased, thus it is not as accurate as LSE since IRF does not consider the possible linear dependency between submatrices in the numerical implementation of WFS. Without realizing the cause of inaccuracy of IRF estimation, Shen and Chung (2006) pointed out that IRF creates less accurate reconstruction by giving an example where the IRF implementation changes the topology of the original surface. We first explore the cause of linear dependence between submatrices of the IRF setting, then we show why LSE and our proposed method give more accurate estimation.

Theoretically, Fourier basis functions are orthonormal. In practical problems, one uses the inner product of the discrete Fourier basis to approximate the theoretical inner product by the definition of the Remannian integral as follows:

$$\langle f_1, f_2 \rangle \;\; = \tfrac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} f_1 f_2 d\mu \;\; \approx \frac{1}{\sum \Delta_i} \sum_{i=1}^{N} f_1(x_i) f_2(x_i) \triangle_i, \tag{12}$$

where $\triangle_i$ is the area element. Therefore the orthonormality of the discrete Fourier basis functions highly depends on the partition of the support of all the basis functions. Since the perfect partition never exists, there is more or less linear dependency between discrete basis functions.

Due to the effects of area elements, the parametrization of the curves and the surfaces can also make the goodness of approximation (12) vary widely. For example, the area-preserving surface parametrization method (Brechbuehler

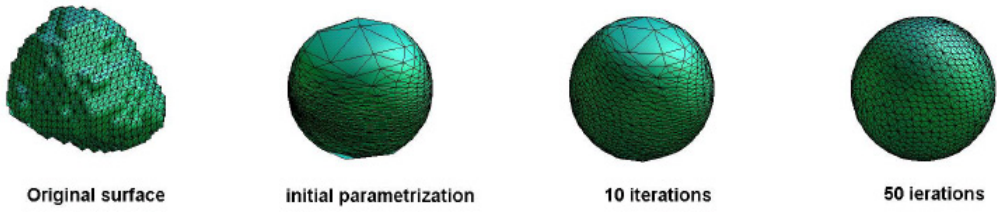

Figure 10: The process of area-preserving parametrization of a given amygdala surface. The original amygdala surface is extracted by Marching-cube method (Lorensen and Cline, 1987). After 50 iterations, the parametrization procedure reaches its tolerance limit and stops.

et al., 1995; Styner et al., 2006) gives nonuniform area elements. Given an area-preserving parametrization, one can check the orthonormality of the Fourier basis generated from this parametrization. We use the inner product matrix of the Fourier basis as

$$M_{in} = (\langle \phi_i, \phi_j \rangle)_{K \times K},$$

where $\{\phi_i\}_{i=1}^K$ are the Fourier basis. Theoretically, if $\{\phi_i\}_{i=1}^K$ are orthonormal, M_{in} should be an identity matrix. In practice, there will always be some noise off the diagonal of M_{in} as shown in Figure 11. We see that with the optimized parametrization (that after 50 iterations), there are still some noises off the diagonal of M_{in}.

We can theoretically explore the reasons and the influence of non-orthogonality on the stepwise regression using a simple example. Let $\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{X}_2 \in \mathbb{R}^2$ as shown in Figure 12. It is clear that $\boldsymbol{X}_1$ is not orthogonal to $\boldsymbol{X}_2$. Using IRF,

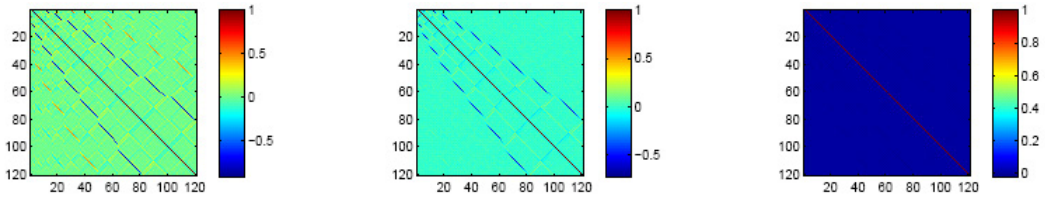

Figure 11: The plots of inner product matrices. The first plot corresponds to the initial parametrization, the second plot corresponds to the parametrization after 10 iterations and the third plot corresponds to the final parametrization after 50 iterations in Figure 10.

one calculates the first residual vector by

$$\boldsymbol{E}_1 \quad = \quad (\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1')\boldsymbol{Y}.$$

The second residual vector is

$$\boldsymbol{E}_2 \quad = \quad (\boldsymbol{I} - \boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2')\boldsymbol{E}_1.$$

But if we use the LSE estimation based on predictor $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, we know that the residual,

$$\boldsymbol{E} \quad = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'})\boldsymbol{Y} \quad = 0,$$

since the space spanned by $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ is $\mathbb{R}^2$ and the projection of $\boldsymbol{Y}$ onto the $(\boldsymbol{X}_1, \boldsymbol{X}_2)$-spanned space is $\boldsymbol{Y}$ itself. $\boldsymbol{E}_2'\boldsymbol{E}_2$ is the variation that can not be explained by the model using IRF.

From Figure 12, $\boldsymbol{E}_1$ is in the subspace spanned by $\boldsymbol{X}_2^*$ since these two vectors are parallel. Therefore

$$\boldsymbol{E}_2^* \quad = \quad (\boldsymbol{I} - \boldsymbol{X}_2^*((\boldsymbol{X}_2^*)'\boldsymbol{X}_2^*)^{-1}(\boldsymbol{X}_2^*)')\boldsymbol{E}_1 = 0.$$

This inspires us to notice that if one replaces $\boldsymbol{X}_2$ with $\boldsymbol{X}_2^*$, then the IRF result will be identical to that of the LSE. Actually one can derive $\boldsymbol{X}_2^*$ from $\boldsymbol{X}_1, \boldsymbol{X}_2$:

$$\boldsymbol{X}_2^* = (\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1')\boldsymbol{X}_2,$$

where $\boldsymbol{X}_2^*$ is the projection of $\boldsymbol{X}_2$ onto the complement of the subspace spanned by $\boldsymbol{X}_1$. One can check the orthogonality,

$$\begin{aligned}
\langle \boldsymbol{X}_1, \boldsymbol{X}_2^* \rangle &= \boldsymbol{X}_1'\boldsymbol{X}_2^* \\
&= \boldsymbol{X}_1'(\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X}_1')\boldsymbol{X}_2 \\
&= \boldsymbol{X}_1'\boldsymbol{X}_2 - (\boldsymbol{X}_1'\boldsymbol{X}_1)(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2 \\
&= \boldsymbol{X}_1'\boldsymbol{X}_2 - \boldsymbol{X}_1'\boldsymbol{X}_2 \\
&= 0,
\end{aligned}$$

which proves that $\boldsymbol{X}_1 \perp \boldsymbol{X}_2^*$. This fact encourages us to carry out extra corrections in the second and later steps of IRF to make all the submatrices orthogonal and thus achieve the same accuracy as LSE. Given a matrix $\boldsymbol{X}$, we denote $SP_{\boldsymbol{X}}$ as the subspace spanned by the columns of $\boldsymbol{X}$, and $P_{\boldsymbol{X}} = \boldsymbol{X(X'X)^{-1}X'}$, the projection matrix of $\boldsymbol{X}$ since $P_{\boldsymbol{X}}\boldsymbol{f}$ gives the projection of $\boldsymbol{f}$ onto $SP_{\boldsymbol{X}}$. We design an adaptive regression algorithm based on the idea of the correction shown in Figure 12:

1. We partition the design matrix into submatrices such that

$$\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_K),$$

where $\{\boldsymbol{Y}_j, j = 1, 2, \cdots, K\}$ are a set of submatrices of $\boldsymbol{Y}$.

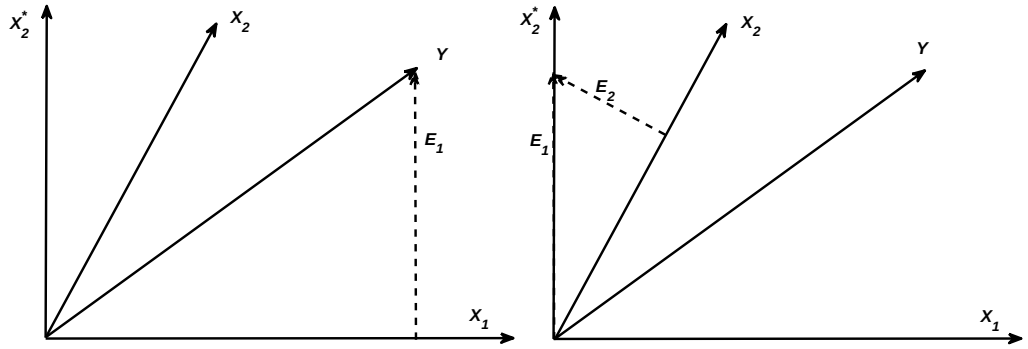

Figure 12: The plots for the example showing why the IRF causes bias. The first plot shows the first step of IRF. The second plot shows the second step of IRF and shows the bias of IRF ($\boldsymbol{E}_2$).

2. We orthogonalize the submatrices using the following procedure:

$$\begin{aligned}
\tilde{\boldsymbol{Y}}_1 &= \boldsymbol{Y}_1 \\
\tilde{\boldsymbol{Y}}_2 &= (\boldsymbol{I} - P_{\tilde{\boldsymbol{Y}}_1})\boldsymbol{Y}_2 \\
\cdots &= \cdots \\
\tilde{\boldsymbol{Y}}_K &= (\boldsymbol{I} - \sum_{j=1}^{K-1} P_{\tilde{\boldsymbol{Y}}_j})\boldsymbol{Y}_K.
\end{aligned}$$

   Note that $\tilde{\boldsymbol{Y}}_i \perp \tilde{\boldsymbol{Y}}_j$, for $1 \leq i \neq j \leq K$.

3. We apply IRF on $\{\tilde{\boldsymbol{Y}}_j\}_{j=1}^K$.

Note that if the dimensions of the submatrices are all 1, the correction step in the new method is exactly the Gram-Schmidt orthonormalization.

Let's denote the residual sequence for IRF as $\{\boldsymbol{e}_j\}_{j=1}^K$, and that for the new method as $\{\tilde{\boldsymbol{e}}_j\}_{j=1}^n$. We also denote the coefficients estimated by the new method

as $\{\tilde{\boldsymbol{\beta}}_j\}_{j=1}^{K}$. We next show

$$\boldsymbol{e}_j'\boldsymbol{e}_j \;\geq\; \tilde{\boldsymbol{e}}_j'\tilde{\boldsymbol{e}}_j, \quad j = 1, 2, \cdots, K,$$

which proves that the new method is more accurate than IRF. By using equation (4.4) of Freund et al. (1961), we have

$$\boldsymbol{\beta}_2 \;=\; P_{\boldsymbol{Y}_2}\boldsymbol{e}_1.$$

Consequently,

$$\boldsymbol{e}_1 = \boldsymbol{Y}_2\boldsymbol{\beta}_2 = P_{\boldsymbol{Y}_2}(\boldsymbol{I} - P_{\boldsymbol{Y}_1})\boldsymbol{Y}_2\tilde{\boldsymbol{\beta}}_2 = P_{\boldsymbol{Y}_2}.\tilde{\boldsymbol{e}}_1$$

We decompose $\tilde{\boldsymbol{e}}_1$:

$$\tilde{\boldsymbol{e}}_1 \;=\; P_{\boldsymbol{Y}_2}\tilde{\boldsymbol{e}}_1 + (\boldsymbol{I} - P_{\boldsymbol{Y}_2})\tilde{\boldsymbol{e}}_1.$$

Hence,

$$\begin{aligned}
\boldsymbol{e}_2'\boldsymbol{e}_2 - \tilde{\boldsymbol{e}}_2'\tilde{\boldsymbol{e}}_2 &= \boldsymbol{e}_1'\boldsymbol{e}_1 - \tilde{\boldsymbol{e}}_1'\tilde{\boldsymbol{e}}_1 \\
&= ((\boldsymbol{I} - P_{\boldsymbol{Y}_2})\tilde{\boldsymbol{e}}_1)'(\boldsymbol{I} - P_{\boldsymbol{Y}_2})\tilde{\boldsymbol{e}}_1 \\
&= \tilde{\boldsymbol{e}}_1'(\boldsymbol{I} - P_{\boldsymbol{Y}_2})\tilde{\boldsymbol{e}}_1.
\end{aligned}$$

This quantifies the difference between two residuals of the second step of IRF and the new method and shows that the new method has a smaller residual. Similarly, we have the difference between two residuals of the third step of IRF and the new method, and so forth. Finally, we have the difference between the final residuals of IRF and the new method as

$$\boldsymbol{e}_K'\boldsymbol{e}_K - \tilde{\boldsymbol{e}}_K'\tilde{\boldsymbol{e}}_K \;=\; \sum_{j=2}^{K} \tilde{\boldsymbol{e}}_{K-1}'(\boldsymbol{I} - P_{\boldsymbol{Y}_j})\tilde{\boldsymbol{e}}_{K-1} \geq 0. \tag{13}$$

Therefore, the *residual of sum squares* (RSS) of IRF is larger than that of the new method, which means that the new method provides more information and has a better fitting based on the same observation and predictors.

One can see that the difference comes from the non-orthogonality between submatrices. But if $\boldsymbol{Y}_{j_0} \perp \boldsymbol{Y}_{j_1}$, $\forall j_0, j_1$, then

$$(\boldsymbol{I} - P_{\boldsymbol{Y}_{j_0}})\tilde{\boldsymbol{e}}_{j_1} = (\boldsymbol{I} - P_{\boldsymbol{Y}_{j_0}})P_{\boldsymbol{Y}_{j_1}} P_{\boldsymbol{Y}_{j_1}} \tilde{\boldsymbol{\beta}}_{j_1} = 0.$$

This indicates that the equality in (13) holds. Therefore, IRF and the new method are identical if and only if

$$\boldsymbol{Y}_{j_0} \perp \boldsymbol{Y}_{j_1}, \quad \forall 1 \leq j_0 \neq j_1 \leq K.$$

Since the new method completely orthonormalized all the submatrices, we call it a complete adaptive iterative regression (cAIR) method. cAIR avoids calculating the inverse of a large design matrix. When using cAIR, one does not have to read the entire design matrix into the computer's memory. The computation becomes more flexible and reliable. Therefore, the implementation is either free of overflow problems or exempted from the loss of accuracy for numerical approximation of the inverse of a large matrix. But sometimes, cAIR is still time-consuming since it is a complete orthogonalization procedure. The same problem happens to Gram-Schmidt orthogonalization in Yeo (2005), in which the author is trying to carry out a Gram-Schmidt orthogonalization procedure for every SPHARM basis function. Gram-Schmidt orthogonalization is a special case of cAIR when the dimension of the submatrices is exactly 1. By our

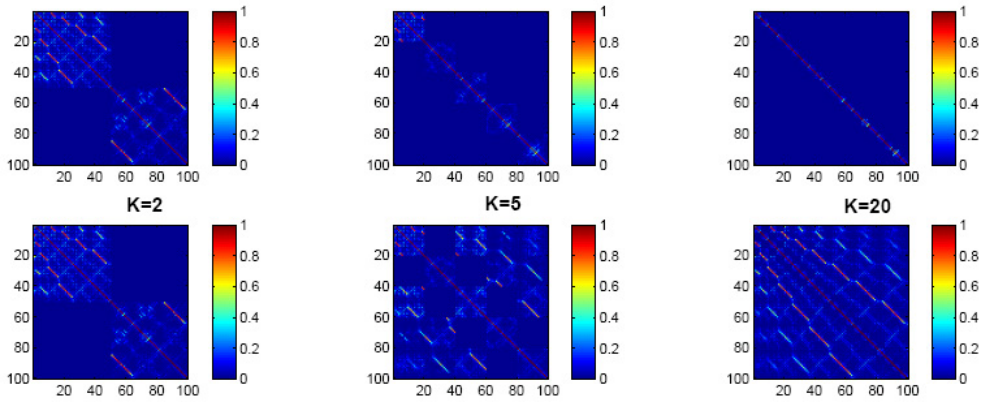

Figure 13: The plots of inner product matrices with corrected design matrices using cAIR and AIR with depth $M = 1$. The first row: the plots of those inner product matrices using cAIR; the second row: the plots of those inner product matrices using AIR. To improve the contrast for the plots, the absolute values of the inner product matrices are used.

experience, one does not have to do a complete orthogonalization. One can only carry out the orthogonalization between neighboring submatrices, and in the meantime, still manage to improve the accuracy. In practice, we design an incomplete *adaptive iterative regression* (AIR), which is trying to eliminate the linear dependence of $M(M \leq K)$ neighboring submatrices to allow a incomplete correction. AIR not only maintains the computational efficiency, but also improves the accuracy. We replace the correction step in cAIR by the following partial correction procedure

$$\begin{aligned} \tilde{\boldsymbol{Y}}_M &= (\boldsymbol{I} - \sum_{j=1}^{M-1} P_{\tilde{\boldsymbol{Y}}_j})\boldsymbol{Y}_K, \\ \tilde{\boldsymbol{Y}}_{M+1} &= (\boldsymbol{I} - \sum_{j=2}^{M-1} P_{\tilde{\boldsymbol{Y}}_j})\boldsymbol{Y}_{M+1}. \end{aligned}$$

We call M the depth of AIR. IRF is a special case of AIR if $M = 0$. The plots of inner product matrices of design matrices, and their corrected counterparts using cAIR and AIR are shown in Figure 13. One can choose the depth M correction of AIR for specific problems. For our experience, $M = 1$ will be sufficient and will be used in the following context.

2.2.3 Automated degree selection using F-statistics

Increasing the degree of WFS will reduces the residuals. But it increases the number of predictors quadratically. Increasing the degree of WFS also increases the risk of over-fitting. Therefore, it is necessary to find the optimal degree that balances the goodness-of-fit and the number of predictors.

In previous Fourier series literatures (Gerig et al., 2001, 2002; Bulow, 2004; Gu et al., 2004; Shen and Chung, 2006), the optimal degree selection has not been addressed. The degrees were simply selected based on a pre-specified error bound that depends on the size of anatomical structure. Even though complex stopping rules exist (for instance, those using GCV and DP), F-statistics are used to determine the stopping rules for stepwise methods since they are easy to implement and have a good intuitive interpretation. One can stop iterations of IRF and AIR when the contribution of certain submatrix is not significant using the hypotheses

$$\begin{aligned} H_0 : &\quad \boldsymbol{\beta}_k = \mathbf{0} \\ H_a : &\quad \text{at least one } \beta_{k,i} \neq 0, \quad i = 1, 2, \cdots, n_k, \end{aligned}$$

where $\boldsymbol{\beta}_k = (\beta_{k,1}, \beta_{k,2}, \cdots, \beta_{k,n_k})$, and n_k is the number of columns of submatrix $\boldsymbol{Y}_k$. Chung et al. (2007b) proposed using the following F-statistic based on the IRF algorithm:

$$F = \frac{(\boldsymbol{e}'_{k-1}\boldsymbol{e}_{k-1} - \boldsymbol{e}'_k\boldsymbol{e}_k)/n_k}{\boldsymbol{e}'_k\boldsymbol{e}_k/(n - \sum_{j=1}^{k} n_j)}. \tag{14}$$

This F-statistic has an intuitive interpretation. The numerator is the improvement in fitting using the last submatrix; the denominator is the estimate variance in response. The F-statistic compares the improvement of each submatrix with the variation of the data.

The same F-statistic of a 2-stepwise regression ($k = 2$ in (14)) was proposed and discussed in Freund et al. (1961); Goldberger (1961) and Alley (1987). Since there is a linear dependency between submatrices $\boldsymbol{Y}_{k-1}$ and $\boldsymbol{Y}_k$, $\boldsymbol{e}_{k-1}$ and $\boldsymbol{e}_k$ are not linearly independent, or $\boldsymbol{e}'_{k-1}\boldsymbol{e}_{k-1} - \boldsymbol{e}'_k\boldsymbol{e}_k$ is not a quadratic form. Therefore it does not have a non-central χ^2-distribution. The linear dependency between submatrices also makes $\boldsymbol{e}'_{k-1}\boldsymbol{e}_{k-1} - \boldsymbol{e}'_k\boldsymbol{e}_k$ and $\boldsymbol{e}'_k\boldsymbol{e}_k$ not statistically independent. As a consequence, the F-statistics for IRF are unlikely to have a non-central F-distribution. Therefore, the comparison of F with the tabulated F-distribution may thus not be very informative for the purpose of assessing significance.

One can also see that the denominator of the test statistic in (14) is always larger than that of AIR and the numerator is always smaller. Therefore, using the F-statistic in (14), for a given threshold $F_{\alpha,n_{k-1},n-(k+1)^2}$ and a significance

level α,

$$P\big(\frac{(\boldsymbol{e}'_{k-1}\boldsymbol{e}_{k-1}-\boldsymbol{e}'_k\boldsymbol{e}_k)/n_k}{\boldsymbol{e}'_k\boldsymbol{e}_k/(n-\sum_{j=1}^k n_j)} \geq F_{\alpha,n_{k-1},n-(k+1)^2}\big)=\alpha,$$

will result in small k. Therefore, IRF is usually conservative in model selection based on the F- statistic in (14).

Similar F-statistic can be defined for AIR:

$$\tilde{F} \;=\; \frac{(\tilde{\boldsymbol{e}}'_{k-1}\tilde{\boldsymbol{e}}_{k-1}-\tilde{\boldsymbol{e}}'_k\tilde{\boldsymbol{e}}_k)/n_k}{\tilde{\boldsymbol{e}}'_k\tilde{\boldsymbol{e}}_k/(n-\sum_{j=1}^k n_k)}. \tag{15}$$

Note that, for AIR, $\tilde{\boldsymbol{e}}_{k-1}\perp\tilde{\boldsymbol{e}}_k$. Then by Pythagorean theorem, $\tilde{\boldsymbol{e}}'_{k-1}\tilde{\boldsymbol{e}}_{k-1}-\tilde{\boldsymbol{e}}'_k\tilde{\boldsymbol{e}}_k$ will be a quadratic form. $\tilde{\boldsymbol{e}}'_{k-1}\tilde{\boldsymbol{e}}_{k-1}-\tilde{\boldsymbol{e}}'_k\tilde{\boldsymbol{e}}_k$ and $\tilde{\boldsymbol{e}}'_k\tilde{\boldsymbol{e}}_k$ are statistically independent. Therefore, $\tilde{F}$ will follow a non-central F-distribution with degrees of freedom $(n_{k-1}, n-\sum_{j=1}^k n_k)$.

For each k, we have

$$\frac{(\tilde{\boldsymbol{e}}'_{k-1}\tilde{\boldsymbol{e}}_{k-1}-\tilde{\boldsymbol{e}}'_k\tilde{\boldsymbol{e}}_k)/n_k}{\tilde{\boldsymbol{e}}'_k\tilde{\boldsymbol{e}}_k/(n-\sum_{j=1}^k n_j)} \;\geq\; \frac{(\boldsymbol{e}'_{k-1}\boldsymbol{e}_{k-1}-\boldsymbol{e}'_k\boldsymbol{e}_k)/n_k}{\boldsymbol{e}'_k\boldsymbol{e}_k/(n-\sum_{j=1}^k n_j)}.$$

Let R be the rejection region. It is straight forward to see that the power function of the test based on AIR

$$P(\tilde{F}\in R)\geq P(F\in R),$$

under the alternative hypothesis, where $P(F\in R)$ is the power function of the tests based on IRF. Then we have the following lemma:

Lemma 2.3. *The F-tests based on equation (15) is more powerful than the ones using equation (14).*

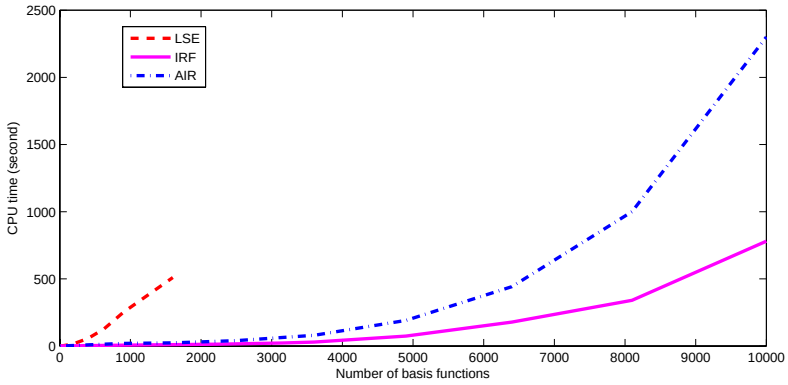

Figure 14: The CPU time of LSE, IRF, AIR representations of a cortical surface with 40962 vertices. The LSE representation met an "out of memory" error with Matlab and stopped if degree is larger than 39 (1600 basis functions). A personal desktop computer with the Pentium 4, 3.2 G Hz CPU and 1 GB memory is used.

2.2.4 Methods comparison

AIR and IRF are specifically designed for large image data. We first assess the capability of the LSE, AIR and IRF representations of large surfaces. A cortical surface (Chung et al., 2006a) with 40962 vertices is used to test the performance of the methods. For this comparison, one only cares about how far (how many basis functions the three methods can use) the three methods can go. We track the CPU time (in the units of seconds) of the three methods for representing the given cortical surface (Figure 14). The experiment is run on a Dell personal computer with Pentium 4, 3.2 G Hz CPU and 1 GB physical memory. LSE ran into an "out of memory" problem in Matlab if one tries to fit a WFS surface with degree larger than 39 (i.e., dimension of its design matrix $> 40962 \times 1600$).

While by using IRF and AIR, one does not have to load the entire design matrix into the memory. We load the 1 submatrix (for IRF) or 2 submatrices (for AIR) at a time into the memory iteratively. By doing this, actually, one can represent the cortical surfaces using AIR and IRF up to arbitrary degrees. In Figure 14, we represent the cortical surfaces by IRF and AIR using up to 10,000 basis functions, but we can definitely go further. By this experiment, we show that one does not have to worry about the problem of loading and computing large matrices using IRF and AIR, which is a real advantage over LSE.

Using the same cortical surface, we also evaluate the efficiency of the F-statistics of IRF and AIR using different bandwidths. As we will see in Figure 15, using the larger bandwidth, IRF and AIR will choose fewer basis functions. For bandwidth $t = 0.1$ and $t = 0.001$, both IRF and AIR give over-smoothed results. The p-value curves of IRF in Figure 15 always go up earlier than those of AIR, which shows that IRF is always a little more conservative than AIR by stopping the iterations earlier and choosing fewer basis functions, even with very well-parameterized data (Chung et al., 2007b). We found that using bandwidth $t = 0.0001$ and 5750 basis functions, AIR seems to give a very good representation of the given cortical surface.

Now we compare the computational efficiency and accuracy of LSE, IRF and AIR methods. We are going to apply three methods to both the simulated data and the amygdala data. From Section 2.2.2, we see that the difference among the three methods is due to the relationship between the submatrices of IRF

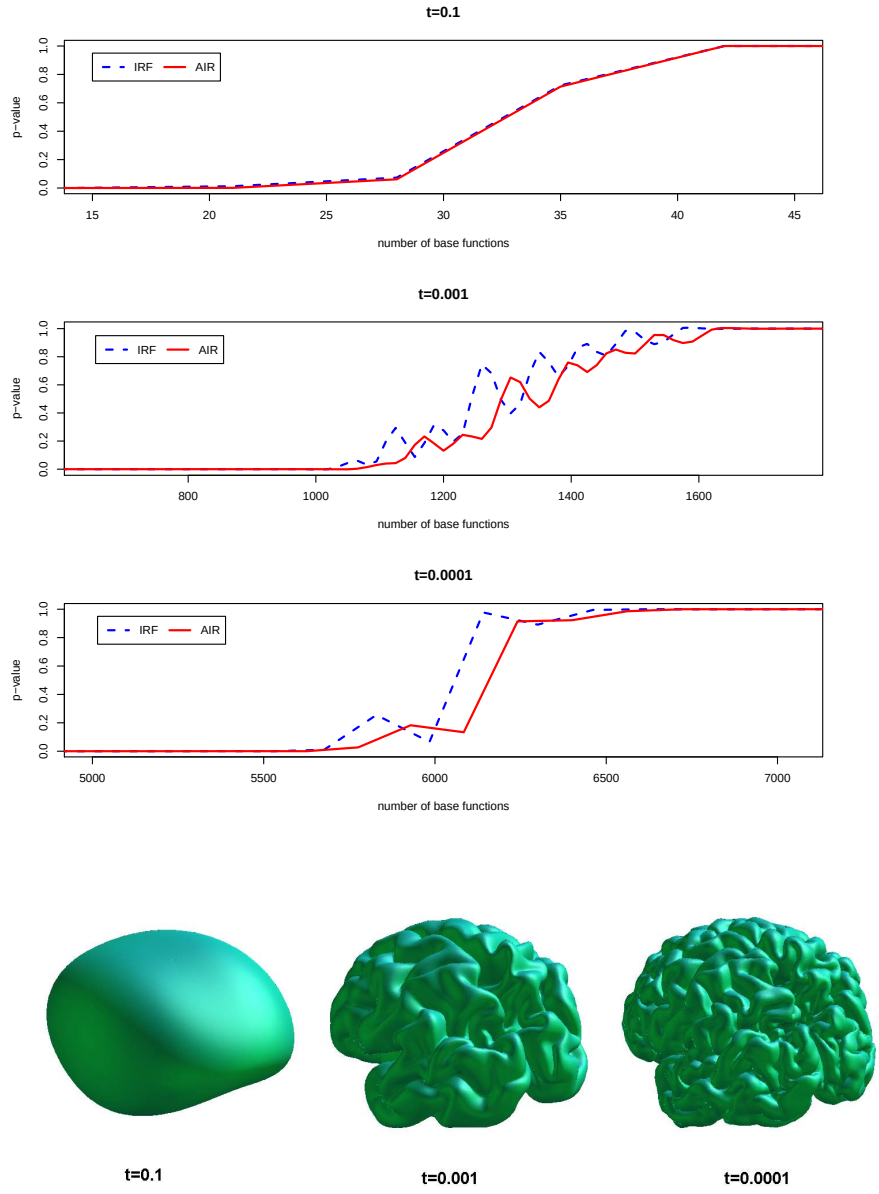

Figure 15: The top 3 rows are the p-value curves using IRF and AIR for bandwidth $t = 0.1, 0.001, 0.0001$. The bottom three cortical surfaces are chosen by AIR for the three pre-specified bandwidths.

and AIR. If the submatrices are not correlated, then the fitted surfaces of the three methods are identical. Therefore, in the simulation, we are interested in the various structures of the design matrices in the related linear models and the correlations between the submatrices generated from the design matrices. We are particularly interested in how different design matrices and their submatrices influence the results and performance of LSE, IRF and AIR. We use *residual sum of squares* (RSS) for a given observation f

$$\text{RSS} = \sum_{i=1}^{n} (\hat{f}_i - f_i)^2$$

and

$$R^2 = \text{cor}^2(\hat{f}, f)$$

to test the goodness of fits, where $\hat{f}$ is the estimation of f. CPU computing times are used to compare the computational efficiency.

In the simulation study, the correlations between the submatrices will be random in order to compare the three methods under different conditions. One should notice that the central idea of this simulation is trying to use different design matrices, since only the variation among design matrices makes the performance of the LSE, IRF and AIR different. The simulation procedure is as follows:

1. A design matrix $\boldsymbol{Y}$ of dimension 2000×240 is randomly generated and fixed.

2. The "true" coefficients $\boldsymbol{\beta}_0$ are given (can also be randomly generated). We assume the true signal

$$\boldsymbol{f} \;=\; \boldsymbol{Y}\boldsymbol{\beta}_0.$$

Our observation is

$$f = E\boldsymbol{f} + \boldsymbol{\epsilon} = \boldsymbol{Y}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$.

3. LSE, IRF, AIR are applied to find the estimation of the signal using the design matrix $\boldsymbol{Y}$ and observation f. For IRF and AIR, one is going to choose the number of submatrices from the set {1, 5, 8, 10, 15, 20, 24, 30, 40, 60, 80, 120, 240} one at a time. In this simulation we assume that all submatrices have the same dimension. The RSS, R^2, and CPU time are saved for each of the three methods.

4. This procedure is repeated for 100 times.

The simulation results are summarized in Figure 16. Note that cAIR and LSE have the same RSS and R^2 values. But the CPU time for cAIR is much higher than those of LSE and IRF and AIR, especially when the numbers of the submatrices are large. When number of submatrices equals 240, the CPU time for cAIR is 5 ± 0.65 seconds, which is much larger than for the other three methods. For a better comparison among LSE, AIR and IRF, we did not include the result of cAIR in Figure 16.

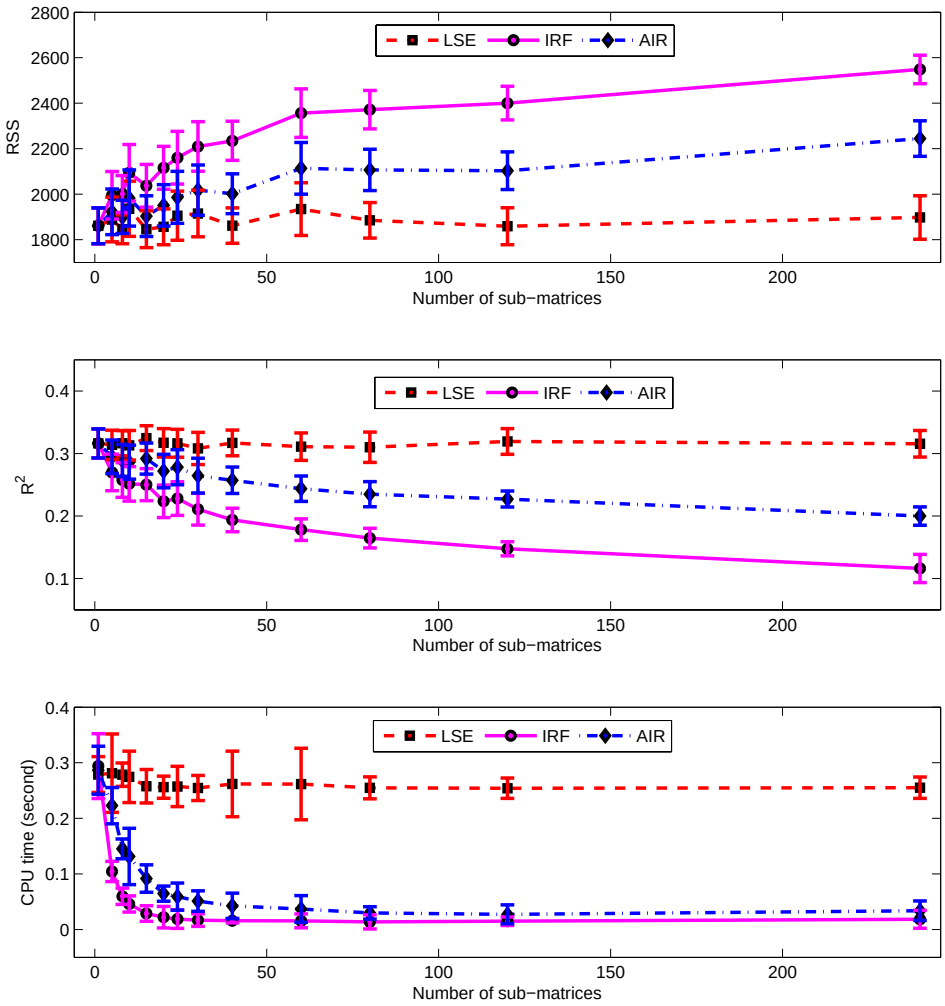

Figure 16: The RSS plot is on the top, R^2 plot is in the middle and CPU time is on the bottom for LSE, IRF and AIR using the simulated data. The curves shows the average values of 100 observations for every number of submatrices from {1, 5, 8, 10, 15, 20, 24, 30, 40, 60, 80, 120, 240}. The error-bars are also added to each curves to show the consistency of the estimation and a rough comparison at each point (number of submatrices).

As we expected, LSE is always the most accurate method with the smallest RSS and the largest R^2-values, which tells us that the LSE estimation provides most information based on the available predictors. AIR's performance on 3 categories is in the middle. The accuracy of AIR is not as good as that of LSE, but is better than IRF. IRF is the fastest method, but with the worst accuracy. The error-bars show the estimated standard errors for estimations. At each point, by viewing the error-bars, one can have a rough idea about what a simple t-test will tell us. For example, from the plot of RSS, the error-bars of the three groups are not overlapped anymore if the number of submatrices is larger than 50, which tells us the difference in the performance of the three methods is significant if using simple t-test. Similarly, the difference in the performance of the three methods is significant if the number of submatrices is larger than 30. If the number of the submatrices is larger than 50, we do not see a significant difference in CPU time between AIR and IRF, even though IRF is slightly faster than AIR.

We also apply the three methods to the amygdala surfaces from the study of autism. CPU time, RSS and R^2 are recorded. The comparison results are summarized in Table 1. The comparison on the amygdala data is similar to that of the simulation study.

From Figure 16 and Table 1, we conclude that IRF is the most computationally efficient and LSE is the least. When the number of submatrices is large, the computational efficiency of AIR is very close to IRF. The order of accuracy

Methods	CPU time ± Std Err	RSS ± Std Err	R^2± Std Err
LSE	16.18 ± 1.24	79.91 ± 13.62	0.997 ± 0.053
IRF	1.33 ± 0.10	160.17 ± 37.45	0.991 ± 0.061
AIR	5.17 ± 0.43	110.52 ± 18.86	0.993 ± 0.058

Table 1: The summary of method comparison of LSE, AIR and IRF on amygdala data of the autism study. the CPU times are in the units of seconds. For every amygdala surface, 256 basis functions are used (up to degree 15 SPHARM basis). For IRF and AIR estimations, each submatrix has 16 columns (so there are 16 submatrices).

of the three methods are LSE, AIR and IRF from the best to the worst.

Chapter 3

Curvature-based Registration

Image registration plays a key role in medical image analysis. It is a process of matching two or more images by minimizing the pre-specified distance between the images. It is a necessary step to remove the translation and orientation difference between images before any comparison and modeling of images could be correctly made. For example, the corpus callosum boundaries are extracted using GVF snakes (Xu and Prince, 1997) as shown in Figure 17. There are both phase and amplitude variations due to the differences of the sizes and positions of the original MR images. The variation is also from the extraction of the corpus callosum boundaries using GVF snakes due to different initialization and image quality. Therefore we need a curve registration procedure to factor out the orientational and translational difference.

One of the major issues of many image registration methods is that it is computationally intensive (Fischer and Modersitzki, 2004). There are various attempts for efficient registrations. Viola and Wells (1995) presented a method based on a formulation of the mutual information between the model and the image using the informative projections of high-dimensional data. Bro-Nielsen and Gramkow (1996) offered a new fast algorithm for non-rigid viscous fluid

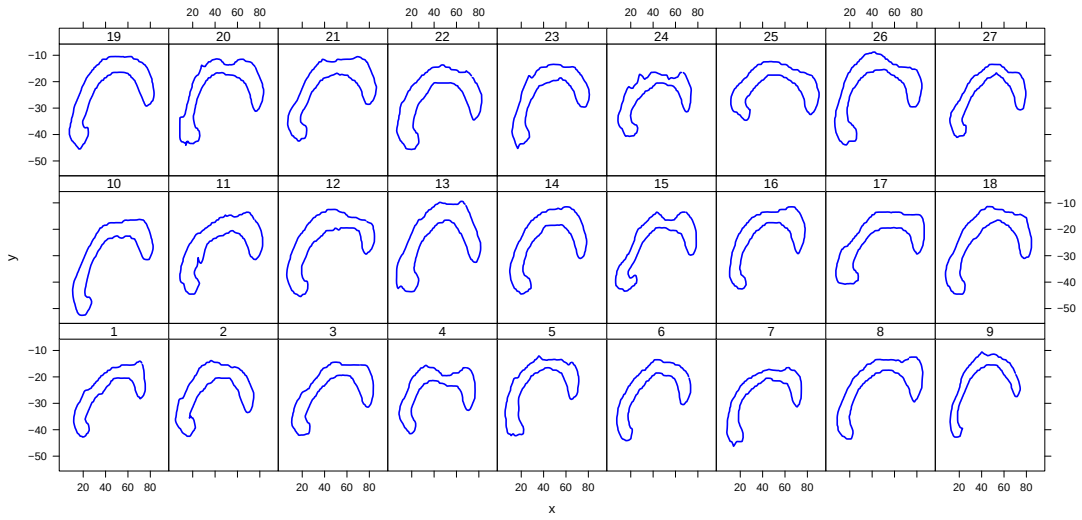

Figure 17: The plots of all the 27 extracted (by GVF snakes (Xu and Prince, 1997)) boundaries of the corpus callosums from the study of autism.

registration of medical images that is based on a linear elastic deformation of the velocity field of the fluid. Fischer and Modersitzki (2004) introduced a new non-linear registration model based on a curvature type smoother. They developed a stable and fast implementation of the new scheme based on a real discrete cosine transformation. One of the key features of these efficient registration schemes is data dimension reduction so that one can represent the data in a parsimonious form, without sacrificing the key features and information of the original data. The data dimension reduction can be done by using the curvature representations. By the first fundamental theorem of plane curves and Bonnet's existence and uniqueness theorem (Stoker, 1969; doCarmo, 1976; Hsiung, 1981; Rubin, 1991), curvature information is independent of locations and rotations

and gives a unique representation of a plane curve or a surface. Curvature functions give a suitable lower dimensional representation. This enables us to design a curvature-based registration method, which is computationally more efficient than those only using coordinates information.

3.1 Curve registration

3.1.1 Curvature estimation

A parametric closed curve $C(s) = (x(s), y(s))$, can be described by two functions, $x(s)$ and $y(s)$. To simplify the closed curves without losing any key feature, we are going to use the curvature functions to represent the corresponding closed curves. The curvature function of a close curve $C(s)$ is defined as

$$k(s) = \frac{x'(s)y''(s) - x''(s)y'(s)}{((x'(s))^2 + (y'(s))^2)^{3/2}}. \quad (16)$$

If $C(s)$ is an arc-length parameterized curve, then $(x'(s))^2 + (y'(s))^2 = 1$. Equation (16) can be simplified as

$$k(s) = x'(s)y''(s) - x''(s)y'(s).$$

By the first fundamental theorem of plane curves, two curves with the same curvature only differ on a rigid-body motion. The corresponding closed curve can be reconstructed from the curvature function by

$$\begin{cases} x(s) = x(s_1) + \int_{s_1}^{s} \cos(\theta(s))ds, \\ y(s) = y(s_1) + \int_{s_1}^{s} \sin(\theta(s))ds, \end{cases}$$

where $\theta(s) = \int_{s_1}^{s} k(s)ds$.

In practice, a closed curve can be represented as a set of ordered points around the curve, where the first and the last points are identical. Let $\{p_i\}_{i=1}^{n}$ be a discrete closed curve. In previous studies (Coxter, 1969; Kreyszig, 1991; Casey, 1996; Gray, 1997; McKeague, 2005), the finite difference methods were used to estimate the underlying curvature functions using equation (16), where the first and second derivatives were approximated by:

$$\begin{aligned} x_i'(s) &= \frac{x_{i+1} - x_i}{s_{i+1} - s_i}, \\ y_i'(s) &= \frac{y_{i+1} - y_i}{s_{i+1} - s_i}, \\ x_i''(s) &= \frac{x_{i+1} - 2x_i + x_{i-1}}{(s_{i+1} - s_i)^2}, \\ y_i''(s) &= \frac{y_{i+1} - 2y_i + y_{i-1}}{(s_{i+1} - s_i)^2}. \end{aligned}$$

A parametrization $\{s_i\}_{i=1}^{n}$ is necessary for the calculation of the first and second derivatives. A natural choice of the parametrization of the given curve (Coxter, 1969; Kreyszig, 1991; Casey, 1996; Gray, 1997; McKeague, 2005) is:

$$s_i = s_{i-1} + \|p_i - p_{i-1}\|, \quad i = 2, 3, \cdots, n, \tag{17}$$

where $s_1 = 0$. Therefore, the finite difference method of the curvature estimation highly depends on the parametrization of the closed curves, which can introduce extra errors to the estimation.

We propose a curvature estimation method, which is independent of curve

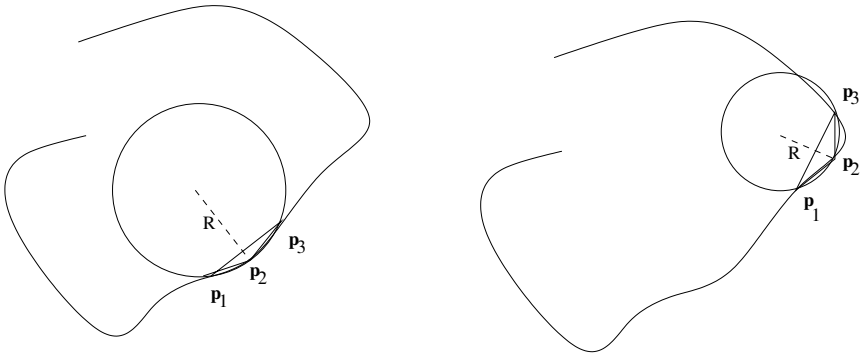

Figure 18: The plots shows the intuition of calculation of curvatures based on the radius of the circle through three consecutive points. $1/R$ is the curvature at point P_2 for both cases. The left plot shows the case where (18) gives very good approximation of the curvature since all the three points are ideally located and spaced. The right plot shows the case that the three point are not ideally located and spaced, the estimation could be a little bit off the true value.

parametrization. The curvature at p_i is calculated as

$$k_i = \text{sign} \cdot \frac{4A(p_{i-1}, p_i, p_{i+1})}{\|p_{i-1} - p_i\| \cdot \|p_{i+1} - p_i\| \cdot \|p_{i+1} - p_{i-1}\|} \tag{18}$$

where $A(p_{i-1}, p_i, p_{i+1})$ is the area of triangle with vertices p_{i-1}, p_i, p_{i+1} and "sign" is 1 if the triangle is inside the closed curve and -1 otherwise. This method is fairly intuitive. The curvature k_i is defined as the inverse of the radius of the circle going through this point and its two neighboring points as shown in Figure 18. It is fairly straight-forward to prove that the estimated curvature using (16) converges to the true underlying curvature.

Theorem 3.1. *Suppose that the second derivative of a closed curve $C(s)$ is continuous at p_i. The underlying curvature of $C(s)$ at p_i*

$$k(p_i) = \lim_{p_{i+1}, p_{i-1} \to p_i} sign \cdot \frac{4A(p_{i-1}, p_i, p_{i+1})}{\|p_{i-1} - p_i\| \cdot \|p_{i+1} - p_i\| \cdot \|p_{i+1} - p_{i-1}\|}.$$

Proof. Let θ be the angle between $p_{i-1} - p_i$ and $p_{i+1} - p_i$ (same angle can be defined between $p_1 - p_2$ and $p_3 - p_2$ in Figure 18). Gonzalez and Maddocks (1996); Wang (2003) showed

$$4A(p_{i-1}, p_i, p_{i+1}) = \frac{1}{2}|p_{i-1} - p_i||p_{i+1} - p_i||\sin\theta|.$$

This equation shows the intuitive connection between the radius of the circle going through the triangle vertices and the standard sine value of θ from elementary geometry. Therefore,

$$\lim_{p_{i+1}, p_{i-1} \to p_i} \text{sign} \cdot \frac{4A(p_{i-1}, p_i, p_{i+1})}{\|p_{i-1} - p_i\| \cdot \|p_{i+1} - p_i\| \cdot \|p_{i+1} - p_{i-1}\|} = \lim_{p_{i+1}, p_{i-1} \to p_i} \frac{p_{i+1} - p_{i-1}}{2\sin\theta}.$$

Let $r(p_{i-1}, p_i, p_{i+1})$ denote the radius of circle going through p_{i-1}, p_i, p_{i+1}. Then,

$$\lim_{p_{i+1}, p_{i-1} \to p_i} \frac{p_{i+1} - p_{i-1}}{2\sin\theta} = \lim_{p_{i+1}, p_{i-1} \to p_i} \frac{1}{r(p_{i-1}, p_i, p_{i+1})}.$$

We finish the proof by the definition of curvature. □

To assess the efficacy of curvature estimation using (18), we introduce a class of closed curves: hypotrochoids (Lockwood, 1961; Lawrence, 1972). A hypotrochoid is determined by three parameters a, b, and h:

$$\begin{cases} x(s) &= (a-b)\cos s + h\cos(\frac{a-b}{b}s), \\ y(s) &= (a-b)\sin s - h\sin(\frac{a-b}{b}s). \end{cases} \tag{19}$$

The class of hypotrochoids includes a variety of curves (see Figure 19). The hypotrochoid curvature function has a closed form:

$$k(s) = \frac{b^3 - (a-b)h^2 + (a-2b)bh\cos(as/b)}{|a-b|(b^2 + h^2 - 2bh\cos(as/b))^{3/2}}.$$

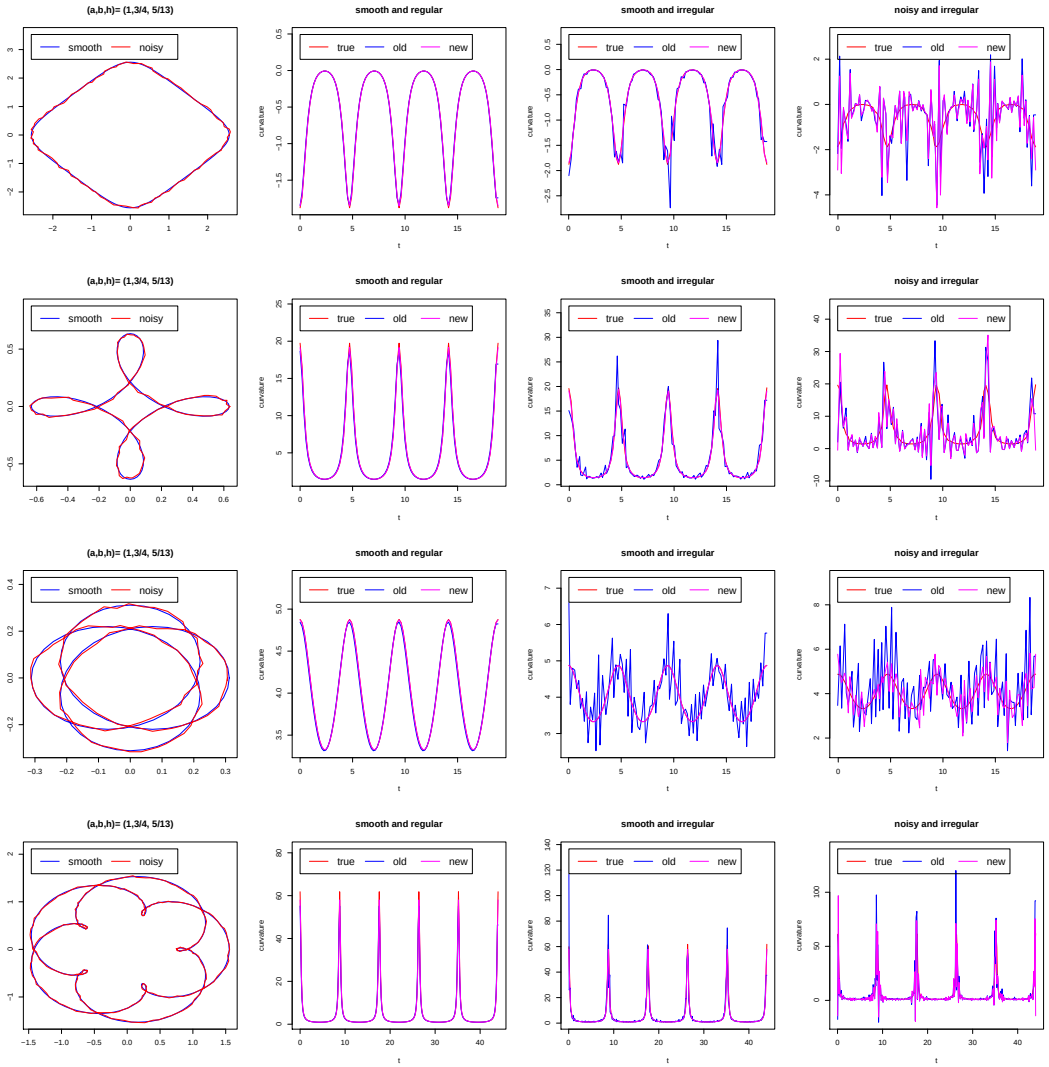

Figure 19: The plots of curvature estimations of 4 special hypotrochoids. The first column is the plots of smoothed or noisy hypotrochoids; the second column is the plots of estimated curvatures of smooth and regularly-spaced curves; the third column is the plots of estimated curvatures of smooth but irregularly-spaced curves; the last column is plots of estimated curvatures of the noisy and irregularly-spaced curves. In the legend, "old" indicates the finite difference method and the "new" indicates our proposed method.

Therefore, the ground truth of hypotrochoid curvatures is always known, which makes it appropriate for assessing the proposed methods of curvature estimation.

For every simulation, three types of hypotrochoids are used to evaluate the proposed curvature estimation method (18): the smooth hypotrochoids with regularly-spaced t's that are calculated directly using (19), the smooth hypotrochoids with irregularly-spaced t's and noisy hypotrochoids with irregularly-spaced t's. The last two types of curves are closer to the real curves obtained in medical image analysis. The results of one simulation are shown in Figure 19. For each hypotrochoid, the true curvature functions, the estimated curvature functions using the finite difference method and the estimated curvature functions using our proposed method are also shown in Figure 19.

Figure 19 shows that our method is clearly better than the finite difference method in curvature estimation for some cases. For the other cases, it is hard to tell the difference. To characterize the goodness of curvature estimation, we use an L^2-norm of the difference between estimated curvature $\hat{k}$ and the true curvature k as

$$\|\hat{k} - k\|_2 \;=\; \sqrt{\int_\Omega (\hat{k}(s) - k(s))^2 ds},$$

where Ω is the range of parameter s.

We repeat the simulation one hundred times. We record all the L^2-norms. The boxplots of the L^2-norms are shown in Figure 20, which shows that our proposed method gives more accurate estimations (with smaller means in the

boxplots) and more robust (with smaller variance) than the finite difference based method.

3.1.2 Curvature-based curve registration

For curve registration, one usually minimizes a pre-specified target functional of the given curve and a template curve (Silverman, 1995; Ramsay and Li, 1997). The WFS representations of the curvature functions are given as $\{k_i(s)\}_{i=1}^{27}$. To estimate the registered curvature functions using a dynamically adjusted template function, one can apply global shift registration method (Ramsay and Silverman, 1997, 2002), in which one is trying to find $\{k_i^*(s)\}_{i=1}^{27}$ that minimizes registration sum squares of errors:

$$\begin{aligned} \text{REGSSE} &= \sum_{i=1}^{27} \int_0^{2\pi} [k_i(s+\delta_i) - \hat{\mu}(t)]^2 ds \\ &= \sum_{i=1}^{27} \int_0^{2\pi} [k_i^*(s) - \hat{\mu}(t)]^2 ds \end{aligned}$$

where the dynamically adjusted template $\hat{\mu}(t)$ is the mean curve of $\{k_i^*(s)\}_{i=1}^{27}$. Therefore, our measure of curve registration is the global sum of squared vertical discrepancies between the shifted curves and the estimated mean curve.

The minimization can be solved iteratively by Newton-Raphson algorithm since $\frac{\partial \text{REGSSE}}{\partial \delta_i}$ and $\frac{\partial^2 \text{REGSSE}}{\partial \delta_i^2}$ have closed forms for this particular case. In practice, the process usually converges within one or two iterations. The registered curvature functions are shown in Figure 21.

The global shift registration does not change the shape of the curvature

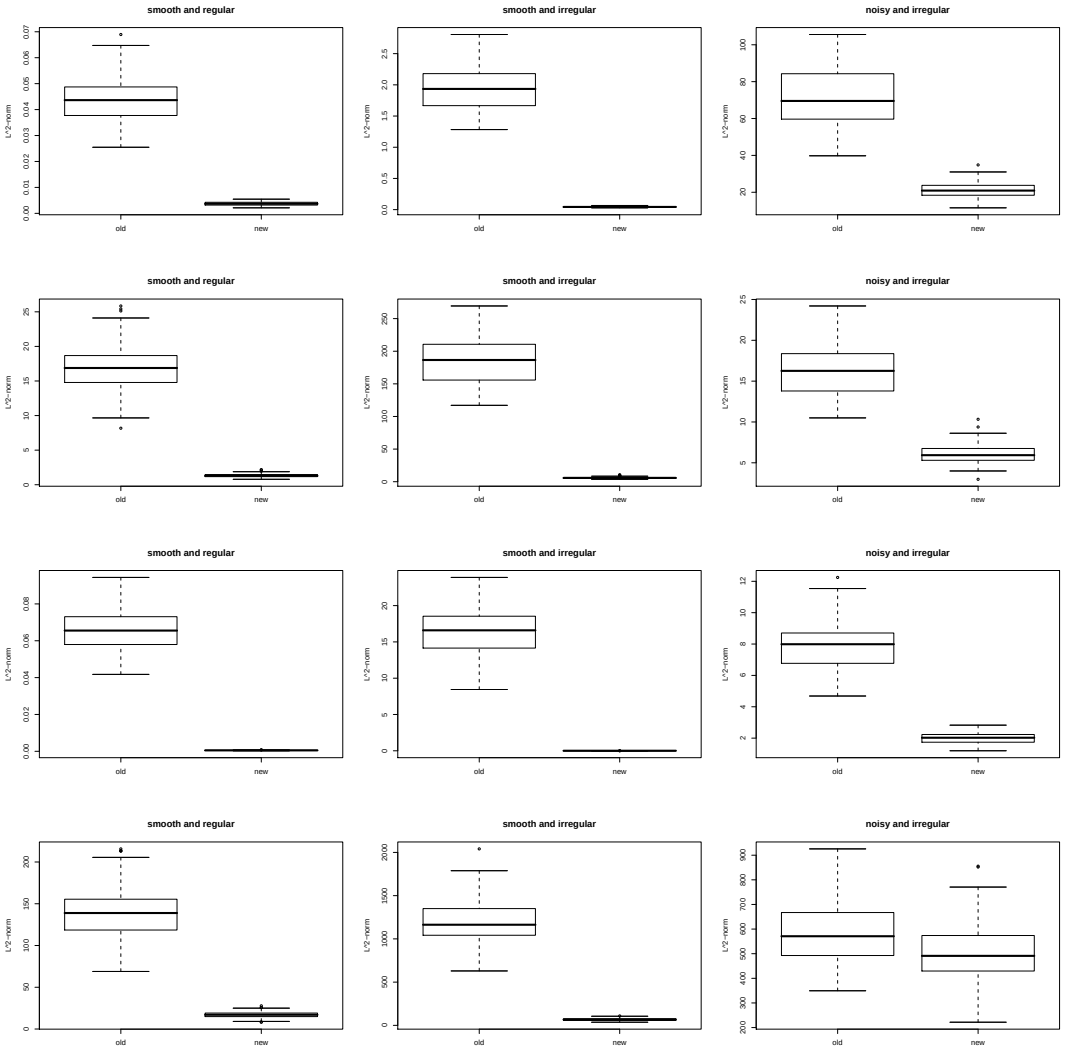

Figure 20: The boxplots of the estimated L^2-norm of the difference between the estimated curvature functions and the true curvature functions. The first column is the boxplots of the L^2-norm of smooth and regularly-spaced curves; The second column is the boxplots of the L^2-norm of smooth and irregularly-spaced curves; The third column is the boxplots of the L^2-norm of noisy and regularly-spaced curves. For the horizontal coordinates, "old" indicates the finite difference method and the "new" indicates our proposed method.

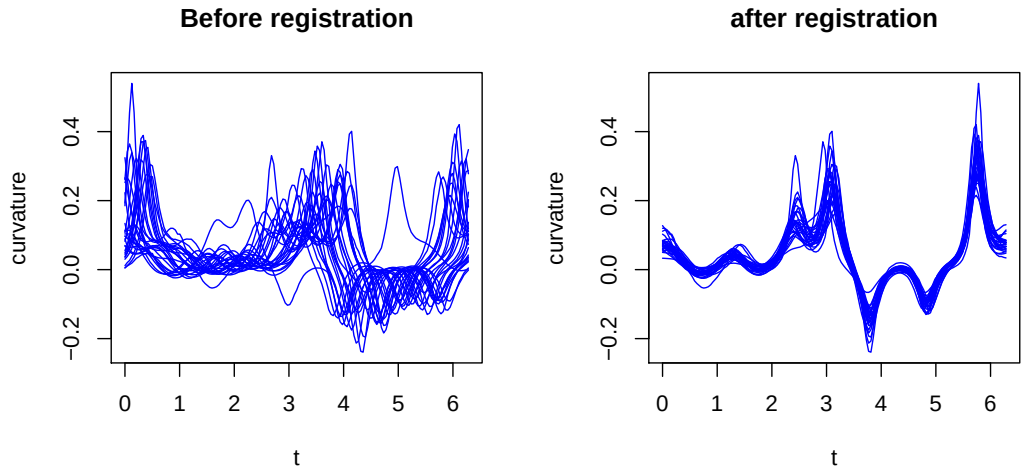

Figure 21: The original curvature functions of 27 GVF snakes (left) and the curvature functions after global shift registration.

function, thus it is equivalent to a global affine alignment. From Figure 21, we see that all the curvature functions are nicely registered. After the global shift registration, the updated cross-sectional average becomes

$$\bar{k}^*(t) = \sum_{i=1}^{27} k_i^*(t).$$

Then we can use this average as a new target (Ramsay and Li, 1997) for computing registered curvature functions. But for the curvature functions in Figure 21, this step seems unnecessary and the improvement is negligible. As pointed out in Ramsay and Li (1997), the curve registration should take place at the level of some derivatives of certain orders rather than the curves itself. Our registration exactly satisfies this criterion. The curvature-based registration not

only reduces the dimensions of the data, but also matches the most important geometric features.

After global shift registration, to further improve the alignment results, one can apply an elastic curve warping method. For a given template curve k_0, we consider the problem of estimating a time-warping function h that minimizes a measure of the penalized L^2-norm

$$V_\lambda = \int \|k_0(s) - k(h(s))\|^2 dt + \lambda \int \frac{(h''(s))^2}{h'^2(s)} ds \tag{20}$$

where h is from a smooth monotone increasing function family and y is the curve-to-be-registered. Similar settings can be found in Ramsay and Li (1997) and McKeague (2005) with minor differences. In (20), h'' controls the smoothness and $1/h'$ prevents $h'(s) = 0$ and therefore controls the monotonicity of the warping function. Therefore, this setting ensures the warping function to be monotone and not too wiggly.

The curvature functions are aligned using elastic warping defined in (20). The alignment results are shown in Figure 22. From the plots, we see that all the curves are almost perfectly aligned. The warping functions are also shown in this Figure. From the warping functions, one can see that most of the variability occurs at the beginning of the curves since the warping functions vary the most at this part. But it seems that the elastic warping does not improve the global shift registration results a a lot. We see the one-to-one mapping of two curves after registration in Figure 23. One can also find all the registered snakes and the mean curves of the autistic and normal control groups after registration in

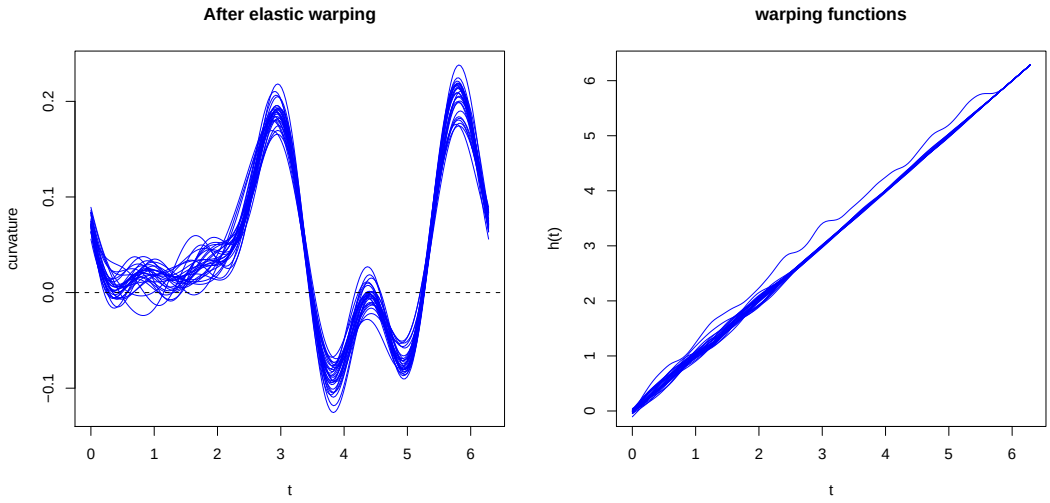

Figure 22: The elastic warping results of the curvatures functions. The warping functions (on the right) are also shown.

Figure 23. The mean plot indicates that there is some difference in the shapes of corpus callosum between autistic and normal control groups.

3.2 Surface registration

Due to the curse of dimensionality, surface registration is always much more complex than curve registration (Audette et al., 2002, 2003), which makes the dimension reduction even more important for surface registration. Similar to the curvature of plane curve, Gaussian and mean curvature are invariant under rigid-body motion for the closed surfaces (Stoker, 1969; Hsiung, 1981). A regular parametric surface can be uniquely reconstructed from the Gaussian and

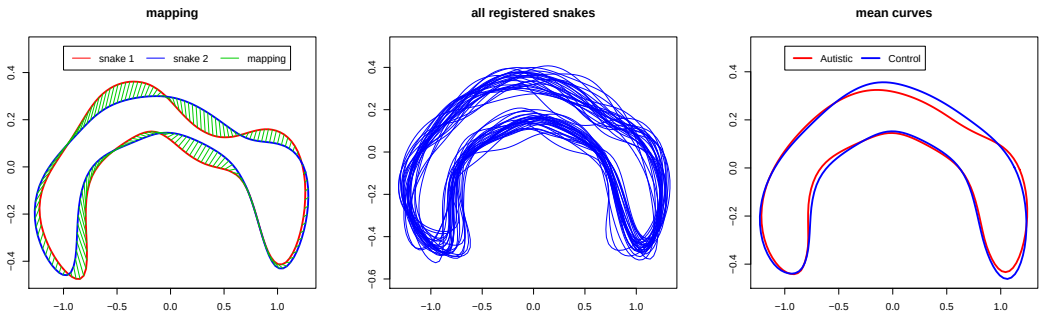

Figure 23: The first plot shows the mapping between two registered snakes; the middle is the plot of all the registered snakes; the last plot shows the mean curves of the autistic and normal control groups.

mean curvatures at each point (Hsiung, 1981; Fan and Nevatia, 1986; Rubin, 1991). Curvatures are frequently used to characterize the local shape of surfaces (Stevens, 1981; Klette and Rosenfeld, 2004; Tong and Tang, 2005). One does not lose any information of the original surface in using only curvatures. The curvature representation of a surface needs only two functions (Gaussian and mean curvature function). But it takes three functions (x, y, and z coordinates) to represent the surface by using the coordinates. Therefore curvature functions give more concise and efficient representations of the surfaces, which makes the surface registration more computationally efficient. WFS representation gives global and analytical forms of the surfaces. This enable us to calculate the Gaussian and mean curvature analytically. We start this section by introducing the first and second fundamental forms of surfaces.

3.2.1 Gaussian and mean curvatures

In spite of the extensive studies and many literatures of surface curvature estimation, results are still not very satisfactory. One of the techniques is to fit a local surface patch and compute partial second derivatives from this patch (Besl and Jain, 1986; Sander and Zucker, 1986; Vemuri et al., 1986; Shi et al., 1994). Derivative computation is very sensitive to noise, therefore it is unstable for real data. Fan and Nevatia (1986) computed the principal curvatures by collecting the four directional curvatures. This method also relies on accurate derivative computation. Shi et al. (1994) fitted a quadric surface locally using the estimated normals. Page et al. (2002) assumed that the surfaces meshes are approximations of piecewise-smooth surfaces derived from range or medical imaging systems. They proposed a normal vector voting algorithm that uses an ensemble of triangles in the geodesic neighborhood of a vertexinstead of its simple umbrella neighborhood to estimate the orientation and curvature of the original surface at that point. Tang (2005) proposed a curvature estimation method based on a local directional curve sampling of the surface where the sampling frequency can be controlled. Unfortunately, the normal estimation requires the surface fitting to be consistent throughout the whole surface. The nature of these algorithm can cause artifact that usually corrupt the output. Piecewise smoothing implies that curvature discontinuities are present where two or more smooth surfaces join, which requires careful consideration. Extra effort has to be made. Using the WFS representation of a surface, one has a

global smooth parametric surface, which makes the derivative estimation robust. One also does not have to worry about the orientation problem of the surfaces associated with the methods that use local fitting. The orientation of a surface is automatically determined by the estimated Gaussian and mean curvatures (Hsiung, 1981).

Many differential geometry textbooks introduce Gaussian and mean curvatures (K, H) using the principle curvatures k_1, k_2 (Stoker, 1969; doCarmo, 1976; Hsiung, 1981; Rubin, 1991; Kuhnel, 2000; Toponogov, 2006):

$$K = k_1 k_2, \quad H = \frac{k_1 + k_2}{2}.$$

But for a parametric surface, Gaussian and mean curvatures are usually explicitly derived from the first and second fundamental forms. Actually, (K, H) are the only invariants of the surface obtained algebraically from the two fundamental forms under rigid-body motion (Hsiung, 1981).

Let $\boldsymbol{r}(\theta, \phi) = (\boldsymbol{x}(\theta, \phi), \boldsymbol{y}(\theta, \phi), \boldsymbol{z}(\theta, \phi))^\tau$ be the WFS representation of a given surface. Define $\boldsymbol{r}_\theta \equiv \partial \boldsymbol{r}/\partial\theta$ and $\boldsymbol{r}_\phi \equiv \partial \boldsymbol{r}/\partial\phi$. The first fundamental form of the surface is written as

$$d\boldsymbol{r}^2 = E d\theta^2 + 2F d\theta d\phi + G d\phi^2. \tag{21}$$

The first fundamental form defines a metric on the surface, therefore, it is also known as the "metric form". E, F and G are also called as Rienmannian metric tensors and the element of area can be defined as (Stoker, 1969; Hsiung, 1981)

$$dS = \sqrt{EG - F^2}.$$

Using the element of area one can compute the total area of the surface as

$$\mathcal{A}(S) = \int_0^{2\pi} \int_0^{\pi} dS \sin\theta d\theta d\phi.$$

The unit normal to the surface can be written as

$$\boldsymbol{n} = \frac{\boldsymbol{r}_\theta \times \boldsymbol{r}_\phi}{\|\boldsymbol{r}_\theta \times \boldsymbol{r}_\phi\|} = \frac{\boldsymbol{r}_\theta \times \boldsymbol{r}_\phi}{dS}. \tag{22}$$

The second fundamental form can be written as

$$d\boldsymbol{r}^2 = e d\theta^2 + 2f d\theta d\phi + g d\phi^2. \tag{23}$$

where

$$e \quad = \boldsymbol{n} \cdot \boldsymbol{r}_{\theta\theta} \quad = -\boldsymbol{n}_\theta \cdot \boldsymbol{r}_\theta, \tag{24}$$

$$f \quad = \boldsymbol{n} \cdot \boldsymbol{r}_{\theta\phi} \quad = -\boldsymbol{n}_\theta \cdot \boldsymbol{r}_\phi, \tag{25}$$

$$g \quad = \boldsymbol{n} \cdot \boldsymbol{r}_{\phi\phi} \quad = -\boldsymbol{n}_\phi \cdot \boldsymbol{r}_\phi, \tag{26}$$

where "$\cdot$" denotes the inner product. The Gaussian curvature K and the mean curvature M can be written based on the first and second fundamental forms

$$K = \frac{eg - f^2}{EG - F^2}, \quad H = \frac{eG - 2fF + Eg}{2(EG - F^2)}. \tag{27}$$

Bonnet's existence and uniqueness theorem for surfaces (Hsiung, 1981; Rubin, 1991) says

Theorem 3.2. *A space surface is uniquely determined by its Gaussian and mean curvatures under rigid-body motion.*

From Theorem 3.2, the Gaussian and mean curvatures represent all the key information of a parametric surface. To calculate the curvatures, one simply needs to estimate the first and the second derivatives of $\boldsymbol{r}$. We start with the estimation of derivatives of WFS representations of general cases.

For the convenience of computation, we simplify the degree K WFS representation of f as,

$$\hat{f} = \sum_{l=0}^{K} \sum_{m=-l}^{l} e^{-\lambda_l t} \beta_{l,m} Y_{l,m} = \sum_{l=0}^{K} \sum_{m=-l}^{l} \alpha_{l,m} Y_{l,m} \tag{28}$$

where λ_l's are the eigenvalues of WFS kernel and $\beta_{l,m}$'s are the coefficients of SPHARM. We start with the derivative of the Legendre polynomials

$$\begin{aligned} \frac{\partial P_l^{|m|}}{\partial \theta} &= \frac{lx P_l^{|m|}(x) - (l+|m|) P_{l-1}^{|m|}}{1-x^2} \\ &= l \cot\theta P_l^{|m|}(x) - \csc\theta (l+|m|) P_{l-1}^{|m|} \end{aligned} \tag{29}$$

where $x = \cos\theta$. Equation (29) is also called the recurrence property of Legendre polynomials. We then derive the derivatives of SPHARM basis recursively

$$\frac{\partial Y_{l,m}}{\partial \theta} = l \cot\theta Y_{l,m} - \csc\theta (l+|m|) Y_{l-1,m} \tag{30}$$

where $Y_{l,m} = 0$ if $m > l$. The derivative of ϕ is relatively easy. We have

$$\frac{\partial Y_{l,m}}{\partial \phi} = -m Y_{l,-m}. \tag{31}$$

Thus

$$\begin{aligned}
\frac{\partial \hat{f}}{\partial \theta} &= (\sum_{l=0}^{K}\sum_{m=-l}^{l} l\alpha_{lm}Y_{l,m})\cdot\cot\theta \\
&\quad +(\sum_{l=0}^{K-1}\sum_{m=-l}^{l}\sqrt{\frac{(2l+3)((l+1)^2-m^2)}{2l+1}}\alpha_{l+1,m}Y_{l,m})\cdot\csc\theta, \qquad (32)\\
\frac{\partial \hat{f}}{\partial \theta} &= \sum_{l=0}^{K}\sum_{m=-l}^{l} m\alpha_{l,-m}Y_{l,m}. \qquad (33)
\end{aligned}$$

Therefore one can compute the first derivative purely based on the coefficients of WFS for given (θ,ϕ). Thus the computation is in general straightforward and fast.

The derivation procedure of the second derivatives of WFS is a little involved. But the formulas turn out not very messy

$$\begin{aligned}
\frac{\partial^2 \hat{f}}{\partial \theta^2} &= -(\sum_{l=0}^{K}\sum_{m=-l}^{l} l\alpha_{l,m}Y_{l,m})\cdot\csc^2\theta+(\sum_{l=0}^{K}\sum_{m=-l}^{l} l^2\alpha_{l,m}Y_{l,m})\cdot\cot^2\theta \\
&\quad -(\sum_{l=0}^{K-1}\sum_{m=-l}^{l} 2(l-1)A^1_{l,m}\alpha_{l+1,m}Y_{l,m})\cdot\csc\theta \\
&\quad +(\sum_{l=0}^{K-2}\sum_{m=-l}^{l} A^2_{l,m}\alpha_{l+2,m}Y_{l,m})\cdot\csc^2\theta, \qquad (34)\\
\frac{\partial^2 \hat{f}}{\partial \theta\partial \phi} &= (\sum_{l=0}^{K}\sum_{m=-l}^{l} lm\alpha_{l,-m}Y_{l,m})\cdot\cot\theta \\
&\quad +(\sum_{l=0}^{K-1}\sum_{m=-l}^{l} mA^1_{l,m}\alpha_{l+1,-m}Y_{l,m})\cdot\csc\theta, \qquad (35)\\
\frac{\partial^2 \hat{f}}{\partial \phi^2} &= -\sum_{l=0}^{K}\sum_{m=-l}^{l} m^2\alpha_{l,m}Y_{l,m}. \qquad (36)
\end{aligned}$$

where

$$\begin{aligned} A^1_{l,m} &= \sqrt{\frac{(2l+3)((l+1)^2-m^2)}{2l+1}}, \\ A^2_{l,m} &= \sqrt{\frac{(2l+5)((l+2)^2-m^2)((l+1)^2-m^2)}{2l+1}}. \end{aligned}$$

The Gaussian and mean curvatures can be computed based on the first and second derivatives explicitly by (27). But one has to pay attention to that there are $1/\sin^2\theta$ terms in the formulas of computing second derivatives, which can cause "being divided by zero" problem in numerical implementation at south and north pole ($\theta = 0$ and $\theta = \pi$) of the parameter space.

The problem of estimating the second derivatives can be avoided. Formulas (24) and (25) tell that we can estimation the second fundamental form via

$$\begin{aligned} e &= -\boldsymbol{n}_\theta \cdot \boldsymbol{r}_\theta, \\ f &= -\boldsymbol{n}_\theta \cdot \boldsymbol{r}_\phi, \\ g &= -\boldsymbol{n}_\phi \cdot \boldsymbol{r}_\phi. \end{aligned}$$

Therefore, to compute the second fundamental form, instead of computing the second derivatives of $\boldsymbol{r}$, we compute the first derivatives of $\boldsymbol{n}$ using the same procedure based on its WFS representation.

To evaluate our proposed curvature estimation method, we use a family of closed surfaces, meta-spheres, which are a generalization of basic harmonic curves and have been used to generate phantoms (vonSeggern, 1994; Xu, 1999).

A meta-sphere $\boldsymbol{r}(\theta,\phi) = (x(\theta,\phi), y(\theta,\phi), z(\theta,\phi))$ is defined as

$$\begin{aligned} x(\theta,\phi) &= (a_1 + b_1\cos(m_1\theta)\cos(n_1\phi))\sin\theta\cos\phi, \\ y(\theta,\phi) &= (a_2 + b_2\cos(m_2\theta)\cos(n_2\phi))\sin\theta\sin\phi, \\ z(\theta,\phi) &= (a_3 + b_3\cos(m_3)\cos(n_3))\cos(\theta). \end{aligned}$$

where $(\theta,\phi) \in [0,\pi]\times[0,2\pi]$, and $\boldsymbol{a} = (a_1,a_2,a_3)$ is the meta-sphere radius in the directions of the three axes, $\boldsymbol{b} = (b_1,b_2,b_3)$ is the ripple amplitude of harmonic components on the meta-sphere, and $\boldsymbol{m} = (m_1,m_2,m_3)$ and $\boldsymbol{n} = (n_1,n_2,n_3)$ are the ripple frequencies. One can also bend the meta-sphere using a simple transformation of the coordinates. For example, a meta-sphere $(x(\theta,\phi), y(\theta,\phi), z(\theta,\phi))$ can be bent in $x-y$ plane as

$$\begin{aligned} \tilde{x} &= x\cos(cx) + y\sin(cx), \\ \tilde{y} &= -x\cos(cx) + y\sin(cx), \\ \tilde{z} &= z, \end{aligned}$$

where c is the parameter that controls the degree of bending. Some sample meta-spheres are shown in Figure 24.

From the definition of meta-sphere, it is conceivable that the computation of the analytical forms of the Gaussian and mean curvature is lengthy and tedious. But with the help of *Mathematica*, one can calculate the first fundamental form

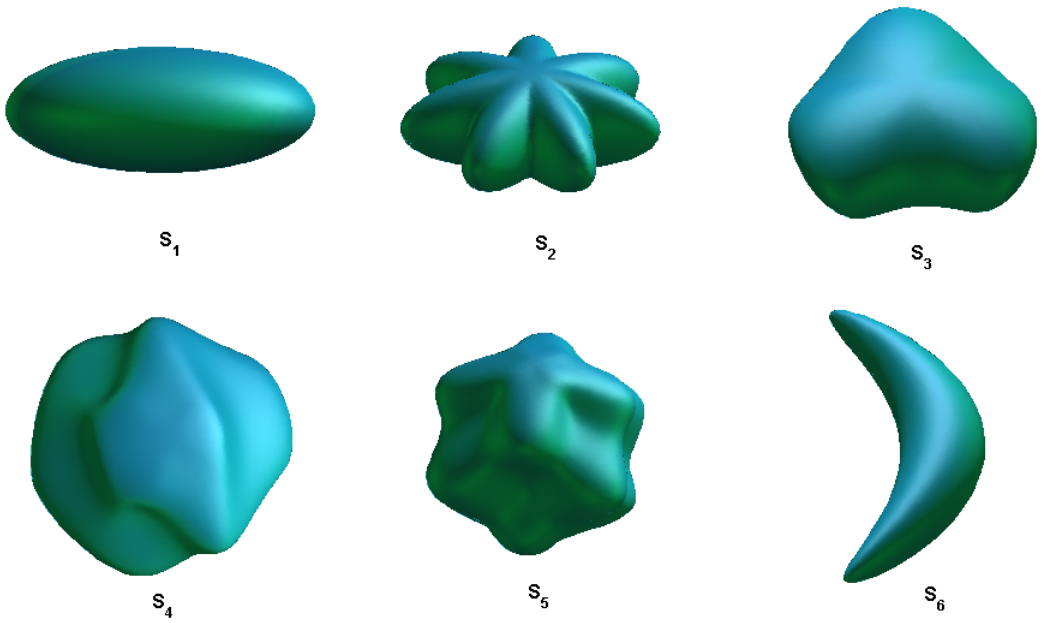

Figure 24: Some sample meta-spheres: $\boldsymbol{S}_1$: $\boldsymbol{a} = (2,3,4), \boldsymbol{b} = \boldsymbol{0}, \boldsymbol{m} = \boldsymbol{0}, \boldsymbol{n} = \boldsymbol{0}, c = 0$; $\boldsymbol{S}_1$: $\boldsymbol{a} = (2,3,4), \boldsymbol{b} = \boldsymbol{0}, \boldsymbol{m} = \boldsymbol{0}, \boldsymbol{n} = \boldsymbol{0}, c = 0$; $\boldsymbol{S}_2$: $\boldsymbol{a} = (2,2,1), \boldsymbol{b} = (0.5,0.5,0), \boldsymbol{m} = (0,0,0), \boldsymbol{n} = (7,7,7), c = 0$; $\boldsymbol{S}_3$: $\boldsymbol{a} = (2,2,1), \boldsymbol{b} = (0.5,0.5,0), \boldsymbol{m} = (0,2,0), \boldsymbol{n} = (3,3,3), c = 0$; $\boldsymbol{S}_4$: $\boldsymbol{a} = (2,2,1), \boldsymbol{b} = (0.5,0.5,0), \boldsymbol{m} = (3,4,3), \boldsymbol{n} = (0,3,0), c = 0$; $\boldsymbol{S}_5$: $\boldsymbol{a} = (2,2,2), \boldsymbol{b} = (0.5,0.5,0), \boldsymbol{m} = (4,4,4), \boldsymbol{n} = (4,4,4), c = 0$; $\boldsymbol{S}_6$: $\boldsymbol{a} = (2,0.5,0.5), \boldsymbol{b} = \boldsymbol{0}, \boldsymbol{m} = \boldsymbol{0}, \boldsymbol{n} = \boldsymbol{0}, c = -0.4$. Some of these 6 meta-spheres are used for validating the curvature estimation method and later used for the registration method evaluation.

precisely as follows

$$
\begin{aligned}
E &= (\cos(\theta)\cos(\phi)(a_1) + b_1\cos(m_1\theta)\cos(n_1\phi)) - b_1 m_1 \cos(\phi)\cos(n_1\phi) \\
&\quad \cdot \sin(\theta)\sin(m_1\theta)^2 + b_3^2 m_3^2 \cos^2(\phi)\cos^2(n_3\phi)\sin^2(m_3\theta) + (\cos(\theta) \\
&\quad \cdot (a_2 + b_2)\cos(m_2\theta)\cos(n_2\phi)\sin(\phi) - b_2 m_2 \cos(n_1\phi)\sin(\theta) \\
&\quad \cdot \sin(m_2\theta)\sin(\phi)^2, \\
F &= (\cos(\theta)\cos(\phi)(a_1 + b_1\cos(m_1\theta)\cos(n_1\phi)) - b_1 m_1 \cos(\theta)\cos(n_\phi) \\
&\quad \cdot \sin(\theta)\sin(m_1\theta))^2 + b_1 m_1 \cos(\phi)\cos(n_1\phi)\sin(\theta)\sin(m_1\theta))^2 + (\cos(\theta) \\
&\quad \cdot (a_2 + b_2\cos(m_2\theta)\cos(n_2\phi)\sin(\phi) - b_2 m_2 \cos(n_2\phi)\sin(\theta)\sin(m_2\theta) \\
&\quad \cdot \sin(\phi))(\cos(\phi)(a_2 b_2 \cos(m_2\theta)\cos(n_2\phi)\sin(\theta) - b_2 n_2 \cos(m_2\theta) \\
&\quad \cdot \sin(\theta)\sin(\phi)\sin(n_1\phi)) + ((a_3 + b_3\cos(m_3\theta)\cos(n_3\phi)\sin(\phi) \\
&\quad -b_3 n_3 \cos(m_3\theta)\cos(\phi)\sin(n_3\phi))^2, \\
G &= ((a_1 + b_1\cos(m_1\theta)\cos(n_1\phi)\sin(\theta)\sin(\phi) + b_1 n_1 \cos(m_1\theta)\cos(\phi) \\
&\quad \cdot \sin(\theta)\sin(n_1\phi)^2 + (\cos(\phi)(a_2 + b_2\cos(m_2\theta)\cos(n_2\phi))\sin(\theta) \\
&\quad -b_2 n_2 \cos(m_2\theta)\sin(\theta)\sin(\phi)\sin(n_2\phi)^2 + (a_3 + b_3\cos(m_3\theta))\sin(\phi) \\
&\quad +b_3 n_3 \cos(m_3\theta)\cos(\phi)\sin(n_3\phi))^2.
\end{aligned}
$$

The second fundamental form can be analogously calculated. Therefore, the ground truth of the meta-sphere curvatures is always known. Then we use our proposed method to estimate Gaussian and mean curvatures of the meta-spheres. The estimated Gaussian and mean curvatures are projected to the (θ, ϕ)-plane for better illustration as shown in Figure 25. One can see, the

estimated curvatures are close to the ground truth. It is hard to tell the difference without very careful examination. But for surfaces it is difficult to put two curvatures in one plot as we have done in the curve curvature estimation. To characterize the difference between the estimated curvatures and the true curvatures, we use the relative errors:

$$100 \times \frac{\hat{K} - K}{K}\% \quad \text{and} \quad 100 \times \frac{\hat{H} - H}{H}\%,$$

where $(\hat{K}, \hat{H})$ is the estimated curvature and (K, H) are the true curvature.

The plot of relative errors are given in Figure 26. There are various patterns of the differences since there is no randomness presented in the two estimation methods. Considering the instability of surface curvature estimation (Besl and Jain, 1986; Sander and Zucker, 1986; Vemuri et al., 1986; Shi et al., 1994), one can find the relative errors of the proposed curvature estimation method are quite small (less than 3%).

3.2.2 Curvature-based affine surface alignment

In this section, we are going to design a curvature-based surface registration. First, Gaussian curvature and mean curvatures of a surface are computed using the estimation of the first and second fundamental forms based on its WFS. Even though it is well-known that a surface can be reconstructed up to second order accuracy if the Gaussian and mean curvatures are known, the reconstruction of the surface using curvature information is very complicated and noise-sensitive (Fan and Nevatia, 1986). The WFS representation of a surface

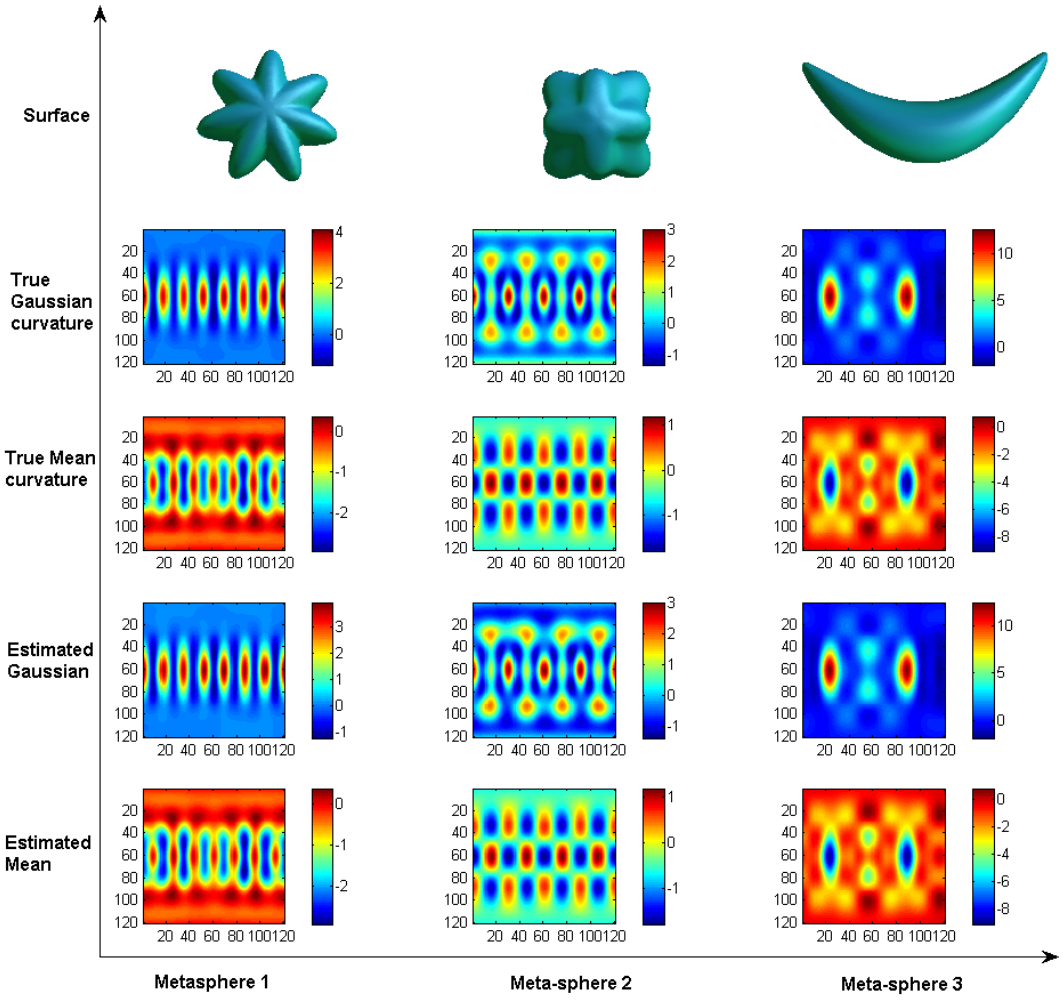

Figure 25: The estimated Gaussian and mean curvatures. The meta-spheres are S_2, S_5 and S_6 in Figure 24. The curvatures are projected onto the (θ, ϕ)-plane. The colors indicate the magnitude of curvatures.

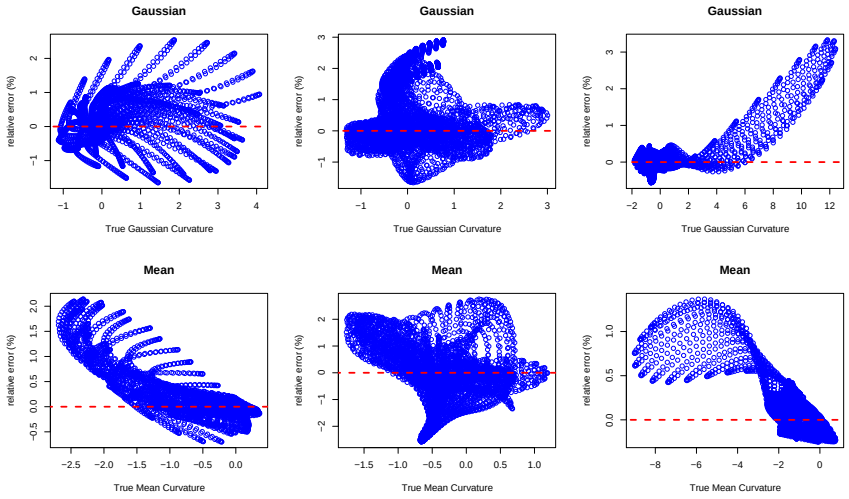

Figure 26: The plots of relative errors of the our proposed curvature estimation method versus true curvature values. The three columns correspond to the three meta-spheres used in Figure 25 respectively.

gives an analytical form on the (θ, ϕ)-parameter space. The Gaussian and mean curvatures share the same parameter space. Then we propose an alignment method purely based on the curvature information. By the one-to-one correspondence of the WFS representation and curvature functions, one can derive the registered surface directly from the registered curvature functions.

A rotation matrix can be generated by three basic rotations about x, y and z-axis. The rotation around the x-axis is defined as:

$$\mathcal{R}_x(\theta_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & \sin\theta_x \\ 0 & -\sin\theta_x & \cos\theta_x \end{bmatrix}$$

where $\theta_x \in [0, \pi]$ is the rotation angle. The rotation matrices are orthonormal

matrix. Therefore, they define a transformation that does not change the size and center location of the surfaces. Similarly the rotations around the y-axis and z-axis are defined as:

$$\mathcal{R}_y(\theta_y) = \begin{bmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{bmatrix}, \mathcal{R}_z(\theta_z) = \begin{bmatrix} \cos\theta_z & \sin\theta_z & 0 \\ -\sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Any 3-dimensional rotation matrix $\mathcal{M} \in \mathbb{R}^{3\times 3}$ can be characterized by the three angles θ_x, θ_y, and θ_z, and may be expressed as a product of 3 basic rotation matrices as

$$M = \mathcal{R}_z(\theta_z) \cdot \mathcal{R}_y(\theta_y) \cdot \mathcal{R}_x(\theta_x).$$

The set of all rotations in $\mathbb{R}^3$, together with the operation of function composition, form the rotation group $SO(3)$.

We can define a transformation matrix that is composed of translation, rotation and scaling as a transformation matrix in the homogenous coordinate system

$$T_{M,t,s} = \begin{bmatrix} s \cdot M_{11} & s \cdot M_{12} & s \cdot M_{13} & t_x \\ s \cdot M_{21} & s \cdot M_{22} & s \cdot M_{23} & t_y \\ s \cdot M_{31} & s \cdot M_{32} & s \cdot M_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{37}$$

where $M = (M_{ij})$ is the rotation matrix, $\boldsymbol{t} = (t_x, t_y, t_z)^\tau$ is the translation vector and s is the scale parameter.

For a given template surface $\boldsymbol{r}_p(\theta,\phi) = (x(\theta,\phi), y(\theta,\phi), z(\theta,\phi))$, the affine alignment of a given surface $\boldsymbol{r}(\theta,\phi)$ is to minimize the L^2-distance between the two surfaces:

$$\arg\min_{M,\boldsymbol{t},s} \int_0^{2\pi}\int_0^{\pi} \|\boldsymbol{r}_p - T_{M,\boldsymbol{t},s}(\boldsymbol{r})\|_2^2 \sin(\theta) d\theta d\phi.$$

This alignment will minimize the orientation and translation difference between two normalized surfaces.

The curvature field of a given parametric surface $\boldsymbol{r}$ is defined as

$$\mathcal{C}(\boldsymbol{r})(\theta,\phi) = (K(\theta,\phi), H(\theta,\phi)),$$

where K and H are the Gaussian and mean curvatures. Similarly to equation (37), one can define a 2D transformation matrix in the homogenous coordinate system as

$$T_{M,t,s} = \begin{bmatrix} s\cdot M_{11} & s\cdot M_{12} & t_x \\ s\cdot M_{21} & s\cdot M_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}.$$

We are looking for a transformation to minimize

$$\int_0^{2\pi}\int_0^{\pi} \|\mathcal{C}(\boldsymbol{r}_p) - \mathcal{C}(T_{M,\boldsymbol{t},s}(\boldsymbol{r}))\|_2^2 \sin(\theta) d\theta d\phi.$$

In general, smaller dimension implies faster and less error-prone solution in optimization. Using the curvature representations, the alignment procedure becomes an optimization problem of four parameters (M, s, t_x, t_y) since the two-dimensional rotation matrix M can be determined by one parameter (the rotation angle). The alignment method using the transformation matrix defined

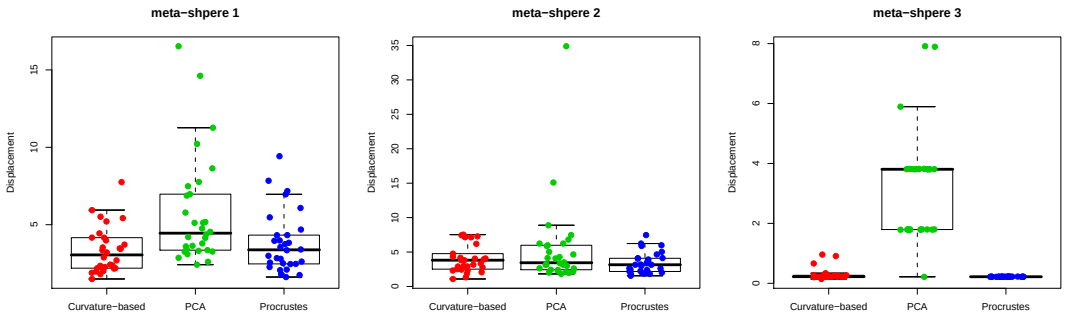

Figure 27: The box-plots of registration scores of the three methods. The jitter plots (colored dots) show the distributions of the registration scores. The three meta-spheres are from Figure 25.

in (37) has seven parameters (s, t_x, t_y, t_z, M), where the three-dimensional rotation matrix M is determined by three rotation angles $(\theta_x, \theta_y, \theta_z)$. Therefore, the surface alignment using the coordinates is an optimization procedure of seven unknown parameters. But the alignment using curvature information is an optimization of four unknown parameters, which shows our proposed curvature-based alignment method is in general more efficient.

We compare our alignment method with PCA alignment method (Shen et al., 2004) and Procrustes alignment (Bookstein, 1997; Styner et al., 2006). The PCA alignment method first computes the three principle components of the surface coordinates, then aligns the surfaces' three principal components of the two surfaces accordingly. Procrustes alignment directly aligns the surfaces to minimize the displacement under rotation, translation and scaling with a set of landmarks. To compare the curvature-based alignment method with these two

Methods	meta-sphere 1	meta-sphere 2	meta-sphere 3
PCA	5.74 ± 3.49	5.38 ± 6.21	3.16 ± 1.77
Procrustes	3.83 ± 1.96	3.40 ± 1.51	0.22 ± 0.02
Curvature-Based	3.34 ± 1.47	4.03 ± 1.96	0.29 ± 0.19

Table 2: The summary of the displacement of the alignments of PCA, Procrustes and curvature-based methods. The entries of the table are the estimated means ± the standard errors of the displacements from the simulations.

methods, we are going to compare the displacement measures, which is defined as

$$\int_0^{2\pi} \int_0^{\pi} \|\boldsymbol{r}_p - \boldsymbol{r}^*\|^2 \sin\theta d\theta d\phi,$$

where $\boldsymbol{r}_p$ is the target surface and $\boldsymbol{r}^*$ is the aligned surface.

For the method comparison, the meta-spheres in Figure 25 are used. For every given meta-sphere, the other surface is generated from this surface by the pre-specified scaling, rotation and translation of the given meta-sphere. Small normal errors are added to the vertices of the surface without changing the topology of the surfaces. Then we use the three methods to align the generated surface to the original surface (the template). After the alignment, the displacements are recorded. This procedure is repeated 30 times for every meta-sphere.

The simulation results can be seen in Figure 27. The registration displacements of the three methods are summarized in Table 2. For meta-sphere 2, the performances of the three methods are very close, even though the curvature-based method and Procrustes method are slightly better than PCA method. For

the first and third meta-spheres, the performances of curvature-based method and Procrustes method are similar. But the curvature-based registration clearly outperforms PCA registration. The difference in the performance of PCA is caused by the fact that PCA does not recognize the directions of the principle components. If the surface is symmetric (like meta-sphere 1), then PCA performs better; otherwise, PCA registration can be very bad and should not be recommended.

Chapter 4

Fast Weighted Fourier Analysis

In Chapter 2 and 3, we have built the systematic ground work of weighted Fourier analysis. In this chapter, we are going to introduce an alternative to weighted Fourier analysis: the fast weighted Fourier analysis, which is closely related to weighted Fourier analysis but approaches the problem from a different angle by using the fast Fourier transforms (FFT).

Model selection (variable selection in regression is a special case) is a bias versus variance trade-off and this is the statistical principle of parsimony (Burnham and Anderson, 1998; Forster, 2000). Efficient and accurate estimation of WFS could also be made via a model selection procedure. As we showed in Chapter 2, the computation of the operations of the large design matrices will be very tedious. But it is always required or implicated for Akaike information criterion (AIC) method (Akaike, 1974), Bayesian information criterion (BIC) method (Schwarz, 1978), stepwise regression method (Hocking, 1976), Mallow's C_p method (Mallow, 1973), LASSO (Tibshirani, 1996) and Dantzig model selection (Osborne et al., 2000; Candes and Tao, 2005). It is time-consuming to compute all the possible models and then select the best model from the model pool. In this chapter, we are going to propose a fast Weighted Fourier model

selection method, which is computationally efficient and gives comparable results with other classic model selection methods such as LASSO and Dantzig model selection.

4.1 Fourier transform

Fourier transform, which was first proposed to solve PDEs, such as Laplace, Heat and Wave equations, has many applications in physics (Greengard (1994) gave a good survey of references for the Fourier (spherical) transform in physics), chemistry (Martyna and Berne, 1989) and biology (Miller et al., 1994). In engineering, Fourier transform is essential in understanding how a signal behaves when it passes through filters, amplifiers and communications channels (Chowning, 1973; Brandenburg and Bosi, 1997; Bosi and Goldberg, 2003). Fourier transform can be also used as high-pass, low-pass, and band-pass filters. It can be applied to signal and noise estimation by encoding the time series (Good, 1958; Harris, 1978; Zwicker and Fastl, 1999).

In this dissertation, we focus on the applications of Fourier transform to image analysis. Fourier transform is a natural image processing tool on image representation which is used to decompose an image into its sine and cosine components. Fourier transform has been widely applied to one of most challenging problems in medical imaging: the resampling and reconstruction of various geometries. Matej and Bajla (1990) proposed a hybrid spline-linear interpolation algorithm for the direct Fourier method. They also compared the

computational requirements of the direct Fourier method algorithm which correspond to distinct interpolation schemes for CT and MR tomography, respectively. Schomberg and Timmer (1995) presented a computational method for reconstructing an n-dimensional signal from a sampled version of its Fourier transform by using a novel gridding method. They found that due to the smoothing effect of the convolution, evaluating the convolution of a signal using a Gaussian kernel is much less error prone than merely interpolating on a regular grid. Hawkins (1996) presented an Fourier transform resampling (FTRS) algorithm, which may be viewed as a generalization of the linear coordinate transformations of standard Fourier analysis by projecting point sources at different transverse positions to estimate cutoff frequency. Taguchi et al. (2001) proposed a method for the implementation of Grangeat's algorithm using spherical transform and applied the method to image reconstruction from cone-beam projections. Bronstein et al. (2002) showed an iterative reconstruction framework for diffraction ultrasound tomography. The proposed algorithm makes use of forward nonuniform fast Fourier transform (NUFFT) for iterative Fourier inversion with incorporation of total variation regularization. Lustig et al. (2004) presented a fast and accurate discrete spiral Fourier transform and its inverse. The inverse solves the problem of reconstructing an image from MRI data acquired along a spiral k-space trajectory. Rowe and Logan (2004), Rowe (2005) and Rowe et al. (2007) used Fourier transform to reconstruct signal and noise of fMRI data utilizing the information of phase functions of Fourier transform

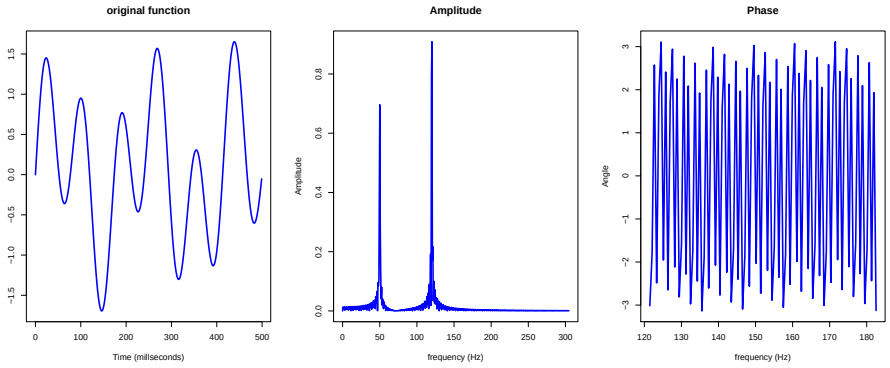

Figure 28: The amplitude (middle) and phase function (right) of the Fourier transform of $g = 0.7\sin(3x) + 0.5\sin(18x)$ on the left.

of images.

Fourier transform has been well-established in Mathematics. As a generalization of Fourier series, the Fourier transform is a linear operator that maps a function space to another function space and decomposes a function into another function of its frequency components. The definition of Fourier transform varies according to different authors (Arfken, 1985; Bracewell, 1999; Krantz, 1999; Trott, 2004). The different definitions are essentially identical with different scaling factors. We are using the routine in Bracewell (1999). Suppose $g \in L(\mathbb{C}), \mathbb{C} = \{x + yi : x, y \in \mathbb{R}\}$. Fourier transform is a linear operator $F : L(\mathbb{C}) \to L(\mathbb{C})$ defined as

$$G(w) = Fg(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(t)e^{-iwt}dt, \quad w \in \mathbb{R}.$$

If g is sufficiently smooth, it can be reconstructed from its Fourier transform

using the inverse Fourier transform

$$g(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(w) e^{iwt} dw.$$

The existence of inverse Fourier transform tells us that a function can be uniquely represented by its Fourier transform. For the purpose of interpretation and visualization, Fourier transform $G(w)$ is usually expressed in polar coordinate as $G(w) = A(w) \cdot e^{ip(w)}$, where $A(w) = \|G(w)\|$ is the amplitude function and $p(w) = \angle G(w)$ is the phase function (as shown in Figure 28).

The Fourier transform on the unit sphere S^2 is also called spherical transform. The spherical transform projects $f \in L^2(S^2)$ into the space spanned by spherical harmonics

$$f(\theta, \phi) \quad = \quad \sum_{l \geq 0} \sum_{\|m\| \leq l} f_{lm} Y_l^m(\theta, \phi) \quad (\theta, \phi) \in [0, \pi] \times [0, 2\pi], \tag{38}$$

where

$$Y_l^m(\theta, \phi) = k_{lm} P_l^m(cos\theta) e^{im\phi},$$

where P_l^m is the associated Legendre function of degree l and order m and k_{lm} is the normalization constant. Here the presentation of spherical transform is different from SPHARM presentation in previous chapters. But they are equivalent as we are going to show later in this chapter.

4.2 Fast Fourier transform

Let observations $\{x_i\}_{i=0}^{N-1}$ be complex numbers. The discrete Fourier transform (DFT) is defined as

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} nk}, \quad k = 0, 1, \cdots, N-1.$$

Computing the N sums directly would take $O(N^2)$ arithmetical operations. A fast Fourier transform (FFT) is an efficient algorithm to compute the DFT and gives the same result using only $O(N \log N)$ operations.

FFT, first discovered by Gauss, has been popularized by Cooley and Tukey (Cooley and Tukey, 1965). Cooley-Tukey FFT algorithm first computes the Fourier transform of the even-indexed numbers and that of the odd-indexed numbers:

$$\begin{aligned} X_k &= \sum_{m=0}^{N/2-1} x_{2m} e^{-\frac{2\pi i}{N}(2m)k} + \sum_{m=0}^{N/2-1} x_{2m+1} e^{-\frac{2\pi i}{N}(2m+1)k} \\ &= \begin{cases} E_k + e^{-\frac{2\pi i}{N}k} O_k & \text{if} \quad k < M \\ E_{k-M} - e^{-\frac{2\pi i}{N}(k-M)} O_{k-M} & \text{if} \quad k \geq M \end{cases} \end{aligned}$$

where E_j is the DFT of the even-indexed numbers and O_j is the DFT of the odd-indexed numbers. One then combines these two results to produce the Fourier transform of the whole sequence. This idea can be performed recursively to reduce the computation time to $O(N \log N)$.

The algorithm described above is called the radix-2 decimation-in-time FFT, which is the simplest and most common form of Cooley-Tukey algorithm. One

can also divide the algorithm into a number of transforms, which is a prime factor of N with slightly degraded in computational speed. This method is called the prime-factor FFT algorithm (Good, 1958). Other important FFT algorithms are also available. The Rader-Brenner algorithm (Rader, 1968) is a Cooley-Tukey-like factorization by reducing multiplications at the cost of increased additions and reduced numerical stability. The Bruun's algorithm (Bruun, 1978) is based on an unusual recursive polynomial-factorization approach and is intrinsically less accurate than Cooley-Tukey in the fact of finite numerical precision. Bluestein's algorithm (Bluestein, 1968) computes the DFT of arbitrary sizes (including prime sizes) by re-expressing the DFT as a convolution.

The accuracy and stability of the algorithms vary. There are many controversies and debates for this aspect. In this dissertation, all the FFT algorithms are based on an open library "FFTW" (Frigo and Johnson, 2005), which uses the most widely accepted Cooley-Tukey algorithm. The multi-dimensional FFT is also well-defined and well-developed in this package. As a base package of Linux operating systems, FFTW is a C subroutine library for computing DFT in one or more dimensions. FFTW is performed on a variety of platforms, which shows that FFTW's performance is typically superior to that of other public available FFT softwares, and is even competitive with vendor-tuned codes. We are particularly interested in the FFT on the 2-sphere (Healy et al., 2003), which uses the techniques of multi-dimensional FFT, but improves it by an efficient

algorithm for the computation of discrete Legendre transforms.

The DFT estimation of $f_{l,m}$ in equation (38) is given as

$$\hat{f}_{l,m} = \frac{\sqrt{2\pi}}{2B} \sum_{j=0}^{2B-1} \sum_{k=0}^{2B-1} a_j^{(B)} f(\theta_j, \phi_k) e^{-im\phi_k} P_l^m(\cos\theta_j),$$

where $0 \leq m \leq l < B$. Notice that the direct computation of every $\hat{f}_{l,m}$ requires $O(B^2)$ arithmetic computation time and thus $O(B^4)$ in total.

Similar to 1-dimensional FFT, the more efficient algorithms use a separation of variables approach. One proceeds by first summing over the k index and computing the exponential summations. One may do this efficiently for all m between $-B$ and B (Elliott and Rao, 1982). This computation requires a discrete Legendre transforms, which is defined as

$$\sum_{k=0}^{N-1} [\boldsymbol{s}]_k P_l^m(\cos(\theta_k)) = \langle \boldsymbol{s}, \boldsymbol{P}_{l^m} \rangle,$$

where $\boldsymbol{s}$ is an arbitrary input vector with k^{th} components $[\boldsymbol{s}]_k$ and $\boldsymbol{P}_l^m$ denotes the vector comprised of appropriate samples of the function $P_l^m(\cos\theta)$.

Healy et al. (2003) solved the subproblems recursively, by further subdivision. Then they combined their solutions to solve the original problem. The advantage of their approach is that the cost of the smaller subproblems, together with the cost of splitting will be less than the cost of direct approach. To insure that the splitting actually results in subproblems of reduced complexity, the three-term recurrence of Legendre functions (this is similar to the recursive property that we used for computing the derivatives of WFS in Chapter 3) is applied. A smoothing and sub-sampling strategy is applied to insure that only

l samples are needed to compute the inner product with a trigonometric polynomial of degree $l < B$. Then this FFT algorithm requires at most $O(B \log^2 B)$ operations.

4.3 Fast weighted Fourier analysis

Even though Fourier transform and Fast Fourier transform are widely used in the field of image analysis, how to choose the significant frequencies is not well studied. Mezrich (1995) proposed an imaging modality that one can choose the dimension of K-space and therefore choose the proper number of frequencies of the observed signal. Wu et al. (1996) obtained the K-space (where MR images are stored) using so called "short-time Fourier transform magnitude vectors". Lustig et al. (2004) also proposed a fast spiral Fourier transform to effectively choose the K-space. Li and Wilson (1995) proposed Laplacian pyramid method to filter out the high frequencies by using a uni-modal Gaussian-like kernel to convolve with images. The problems with those model selection methods and procedures are that they did not consider the possibility that even some low frequencies are not necessarily significant. They simply picked all the low frequencies using a brutal-force thresholding and threw away the high frequencies.

As mentioned in Chapter 2, the eigenfunctions $\{\phi_j\}_{j=1}^n$ of the Laplacian operator Δ are orthonormal. But for the numerical implementation, the discrete eigenfunctions are only approximately orthonormal if the curve or surface is well-parameterized. To check the orthonormality, we use the inner product matrices

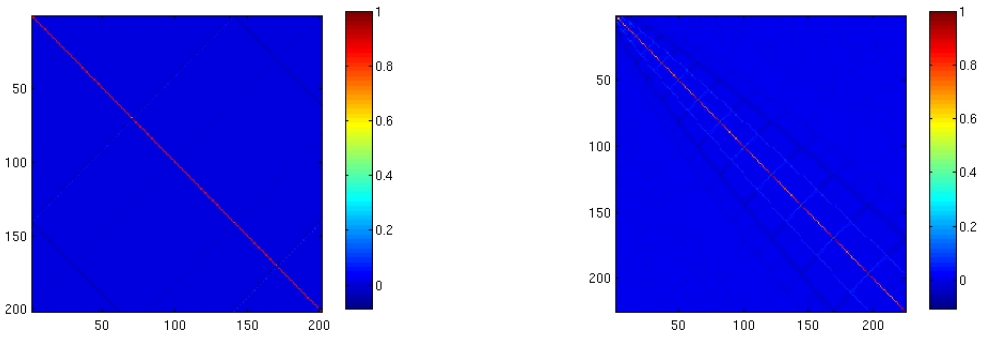

Figure 29: The colormap of inner product matrix of 200 Fourier basis functions based on the parametrization of a GVF snake boundary of the corpus callosum used in the study of autism (left) and colormap of the inner product matrix of 225 (degree 14) SPHARM basis functions based on the parametrization of a amygdala surface.

defined in Chapter 2

$$M = (\langle \phi_i, \phi_j \rangle)$$

where $\{\phi_k\}_{k=1}^{N_1}$ are a set of one-dimensional Fourier series basis functions or a set of SPHARM basis functions. The colormaps of inner product matrices are shown in Figure 29. From the plots, we see that the matrices are dominated by their diagonals. But there are small noises off the diagonals of the matrices, which show that the basis functions are not exactly othornormal.

We are interested in the inverses of the inner product matrices. Actually, it can be proved that their inverse matrices are also dominated by their diagonals.

Lemma 4.1. *Let I be the $n \times n$ identity matrix and J be the matrix with all*

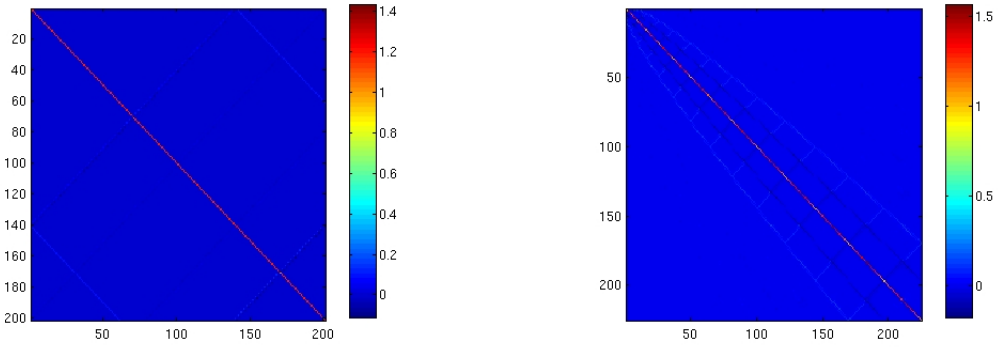

Figure 30: The inverse of colormap of inner product matrix of Fourier basis functions (left) and inverse colormap of that of SPHARM basis functions. The corresponding inner product matrices are shown in Figure 29.

the entries smaller than 1, and $b = o(a)$*. Then we have*

$$(aI + bJ)^{-1} \approx \frac{1}{a}I - \frac{b}{a^2}J. \tag{39}$$

Note that the inner product matrices also have the format of $aI + bJ$, which is dominated by the diagonals. The conclusion can be easily proven from

$$(aI + bJ)(\frac{1}{a}I - \frac{b}{a^2}J) \approx I.$$

The matrix Taylor expansion of $(aI + bJ)^{-1}$ gives the same result. The conclusion can also be easily demonstrated by plotting the inverses of inner product matrices as shown in Figure 30.

We next show that this property of the inner product matrices of the Fourier basis functions is crucial for the fast weighted Fourier analysis. In weighted Fourier analysis, the linear model we used for estimating the coefficients of

WFS is

$$\boldsymbol{f} = \boldsymbol{Y}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I}).$$

The simulations in Chapter 2 show that it is appropriate to assume normality. Using the following lemma, we will establish our proposed model selection procedure.

Lemma 4.2. *Suppose that* $\boldsymbol{f}$ *follows a multivariate normal distribution with mean* $\boldsymbol{Y}\boldsymbol{\beta}$ *and covariance matrix* $\sigma^2\boldsymbol{I}$*, then the LSE of* $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{f} \sim N_p(\boldsymbol{\beta}, (\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\sigma^2). \tag{40}$$

Given that the columns of $\boldsymbol{Y}$ are the Fourier basis functions or SPHARM basis functions, the covariance matrix $\sigma^2(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}$ is exactly the inverse of the inner product matrix of the basis functions. Since $\boldsymbol{Y}^T\boldsymbol{Y}$ is dominated by its diagonal,

$$(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\sigma^2 = c_0 I - d_0 J,$$

where c_0 and d_0 are constants and $d_0 = o(c_0)$. This matrix is also dominated by its diagonal. From Lemma 4.2, we have the marginal distribution,

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2(c_0 - d_0)) \quad i = 1, 2, \cdots, K, \tag{41}$$

where K is the number of the columns of $\boldsymbol{Y}$. We are trying to eliminate the

insignificant β_i's based on the following hypothesis tests

$$H_0: \quad \beta_i = 0,$$

$$H_a: \quad \beta_i \neq 0$$

for $i = 1, 2, \cdots, K$. Based on the result in (41), the test statistic will be the t-statistic

$$T_i = \frac{\hat{\beta}_i}{\text{Std. Err. of } \beta_i} \approx \frac{\hat{\beta}_i}{\hat{\sigma}\sqrt{c_0 - d_0}}.$$

Then $\|T_i\| \geq t_{0.025,n-1}$ gives the threshold at 0.05 significance level

$$\|\beta_i\| \geq b_0 \approx t_{0.025,n-1}\hat{\sigma}\sqrt{c_0 - d_0}.$$

where n is the number of observations.

Therefore, the significant frequencies of WFS can always be chosen using their coefficients by giving a constant threshold. But for WFS based image analysis, estimation of the coefficients is usually time-consuming. One needs to find an alternative and faster way to compute coefficients. In the next two sections, we are going to show that the coefficients can be computed efficiently by fast Fourier transform (FFT).

Therefore, the framework of our model selection method is designed as follows:

1. For a given observation $\boldsymbol{f}$, which is usually a curve or a surface, the Fourier transform of $\boldsymbol{f}$ is computed via FFT.

2. The coefficients of Fourier series are derived from the results of FFT.

3. The covariance matrix of $\boldsymbol{\beta}$ is derived from the first K basis functions using Lemma 4.2 and σ is estimated by

$$\hat{\sigma} = \sqrt{\frac{1}{n-K}\|\boldsymbol{f} - \boldsymbol{Y}\boldsymbol{\beta}\|^2}$$

where $\boldsymbol{Y}$'s columns are the K basis functions and $\boldsymbol{\beta}$ is estimated only using K basis functions. Then the standard error of β_i $(i = 1, \cdots, K)$ is the ith diagonal entry of matrix $(\boldsymbol{Y'Y})^{-1}\hat{\sigma}$

4. The threshold is then

$$b_0 = \lambda t_{0.025,n-1} \cdot \hat{\sigma}\sqrt{1-b},$$

where $\lambda = 1$ is always applied and b is the estimated maximum of the off-diagonal of $\boldsymbol{Y'Y}$. But for more flexibility, λ can be changed accordingly to various conditions to find the suitable results.

5. The frequencies with coefficients larger than the threshold are chosen by the method.

This procedure selects the significant coefficients $\boldsymbol{\beta}_s = (\beta_{1,s}, \beta_{2,s}, \cdots, \beta_{n_s,s})$. Then the final WFS representation is

$$\hat{f} \;=\; \sum_{k=1}^{n_s} e^{-\lambda_{k,s}}\beta_{k,s}\phi_{k,s}$$

where $\lambda_{k,s}$ and $\phi_{k,s}$ are the selected eigenvalues and eigenfunctions (basis functions). We call this model selection procedure as fast weighted Fourier analysis.

4.4 One-dimensional fast weighted Fourier analysis

Most of the applications and generalizations of Fourier transform are based on the following standard properties of Fourier transform:

Lemma 4.3. *For a given bounded continuous integrable function (e.g. f), we denote the corresponding capital letter (e.g. F) as its Fourier transform.*

a. *If $g(x) = f(x-a)$, then $G(w) = e^{-iaw}F(w)$.*

b. *If $g(x) = f(x/\lambda)$, then $G(w) = \lambda F(\lambda w)$.*

c. *If $h = f * g$, the convolution of f and g, then $H(w) = F(w)G(w)$.*

d. *If $d(x) = f'(x)$, then $D(w) = iwF(w)$.*

e. *If $f(x) = \cos(2\pi w_0 x)$, then $F(w) = \delta(w+w_0) + \delta(w-w_0)$; If $f(x) = \sin(2\pi w_0 x)$, then $F(w) = \delta(w+w_0) + \delta(w-w_0)$.*

We derive the Fourier series using the corresponding Fourier transform.

Lemma 4.4. *One-dimensional Fourier series of $f \in L^2(\mathcal{M})$ have the following format*

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos(nx) + b_n \sin(nx))$$

we have

$$\begin{aligned} F(w) &= \frac{a_0}{2}\delta(w) + \sum_{n=1}^{\infty}(a_n(\delta(w+n) - \delta(w-n)) \\ &\quad + b_n(\delta(w+n) - \delta(w-n))) \\ &= \frac{a_0}{2}\delta(w) + \sum_{n=1}^{\infty}((a_n + b_n)\delta(w+n) + (a_n - b_n)\delta(w-n)). \end{aligned} \quad (42)$$

Equation (42) holds using (e) in Lemma 4.3.

In practice, we are trying to estimate signal $g(x), x \in [0, 2\pi]$. Only noisy signal is observed as

$$g_1(x) = g(x) + \epsilon(x),$$

where $\epsilon(x) \sim N(0, \sigma^2)$ is the white noise. One is trying to find the Fourier series representation to approximate the true signal

$$\hat{g} = \frac{a_0}{2} + \sum_{n=1}^{K}(e^{-n^2 t} a_n \cos(nx) + b_n \sin(nx)),$$

where K is selected manually or automatically.

We are going to demonstrate the fast weighted Fouriere analysis methods with simulated data and corpus callosum data. The first simulation is to estimate the sinusoid signals. In this simulation, we let

$$g_1(x) = 0.7\sin(7x) + \sin(18x) + \epsilon,$$

where $\epsilon \sim N(0, 0.2^2)$ as shown in Figure 31.

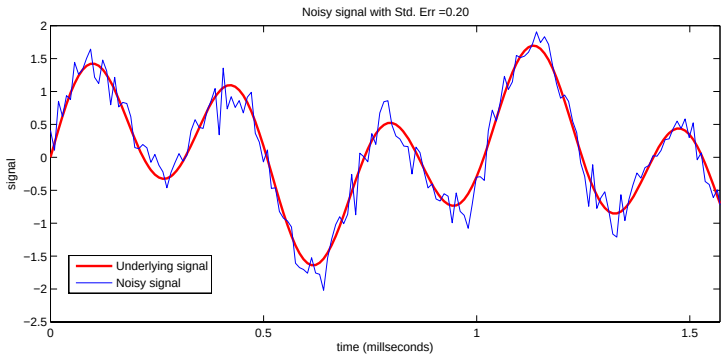

Figure 31: The underlying and noisy curve used in the simulation with true signal $0.7\sin(7x) + \sin(18x)$.

To estimate the signal $g(x)$ using LSE, one has to generate at least $2 \times 18 + 1$ basis functions to capture the high frequency information (of degree 18 Fourier basis functions). It is likely to have the over-fitting problem for LSE using redundant predictors. On the other hand, using fast weighted Fourier analysis, one can easily find that two basis functions are enough for our analysis. It also provides the estimation of coefficients of the corresponding basis functions as shown in Figure 32. When using 1000 observations, the amplitudes are not exactly at 0.7 and 1. The main reason is the presence of noise. The other reason is that one has finite range of observations while the Fourier transform is defined over the whole real line. If one increases the range of observations, as shown in Figure 32, we have a better approximations.

The threshold is computed based on the observations, which is shown as the

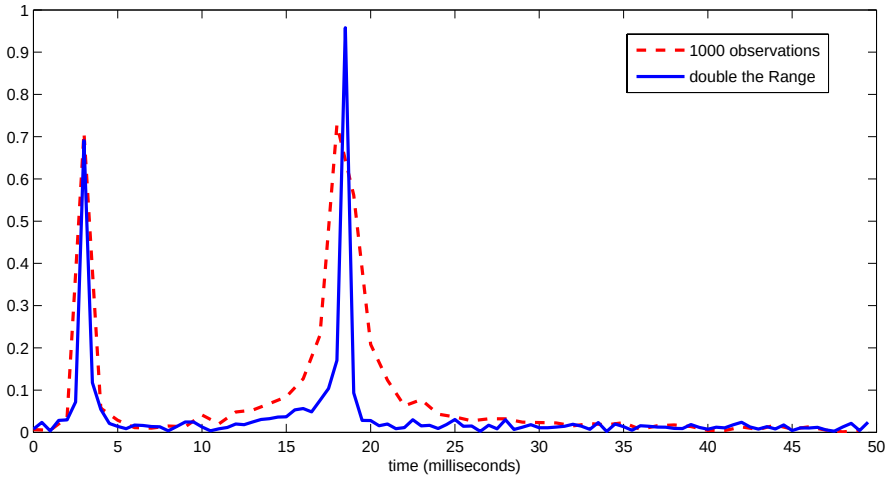

Figure 32: The fast Fourier transform results using different observation ranges. "double the range" means the the support of observed function is doubled.

dashed line in the Figure 32. Using the results of Fourier transform, the estimated signal functions are shown in Figure 33. When using 1000 observations, the estimation is over-smoothed. But if we increase the range of observations, we have a very good estimation of the original signal.

From first simulation, we see that, for the estimation of trigonometric functions or their combinations, Fourier transform will give better and faster results than least-squares estimation. Considering that if the signal function has high frequency component (e.g. the component $\sin(nx)$ when n is very large), the least-squares estimation will be very inefficient and very likely to have overfitting problems using all the $2n + 1$ basis functions.

For the second simulation, we assess the performance of fast weighted Fourier

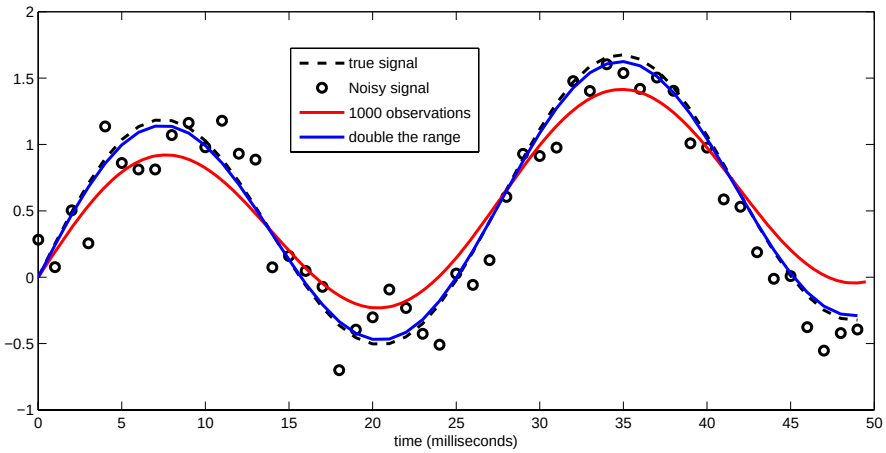

Figure 33: The final result of fast weighted Fourier analysis for the first simulation. Two estimated curves are given: one is using 1000 observations, and the other one is using 2000 observations.

analysis on the estimation of a more general signal. Let the true signal be

$$g(x) = \begin{cases} x^2 \cdot (x - 2\pi)^2, & x \in [0, 2\pi] \\ g(x + 2\pi), & \text{otherwise.} \end{cases}$$

Note that $g(x)$ is periodic and smooth (its first derivative is continuous) as shown in Figure 34. For the general curve that we defined, one still manages to find a good approximation of the true signal as shown in Figure 35.

We can also apply our method to the corpus callosum (CC) data. GVF snakes algorithm will provide noisy boundaries of CC's. So a smooth CC boundary should be achieved for statistical analysis. First, using the arc-length parametrization method (as described in Chapter 2), for each obtained discrete

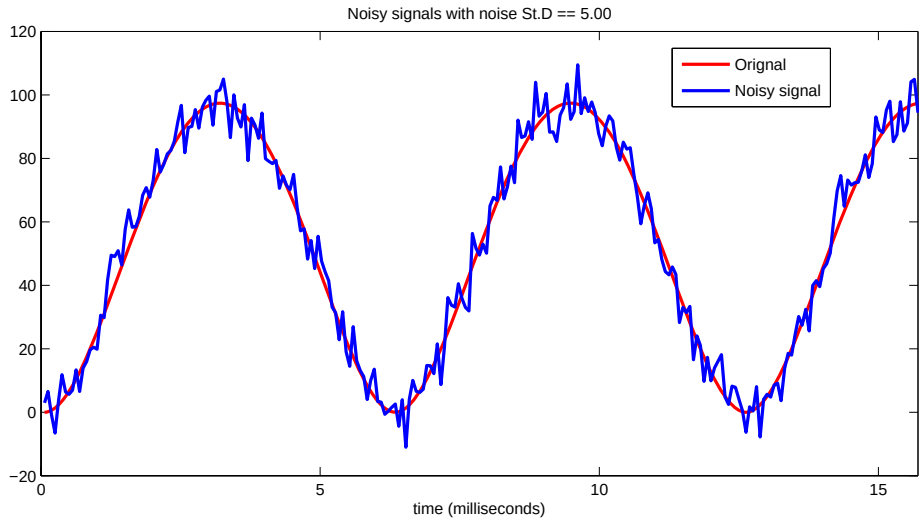

Figure 34: A noisy non-trigonometric curve with underlying true signal $x^2(x - 2\pi)^2$ (the smooth curve).

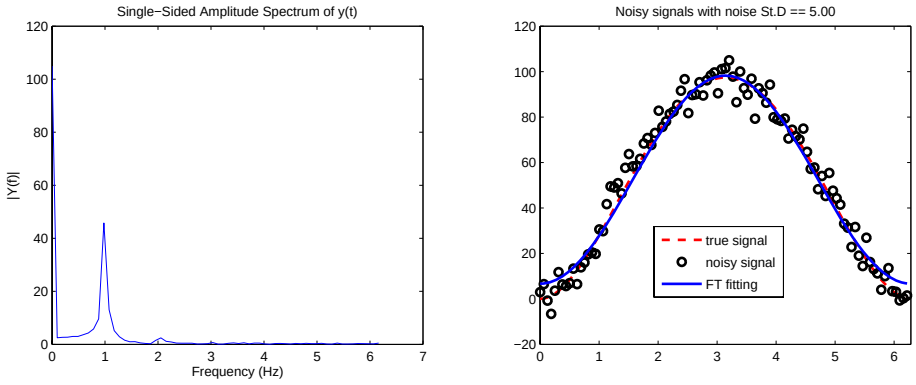

Figure 35: The FFT results (left) and the estimated signal for the observations in Figure 34.

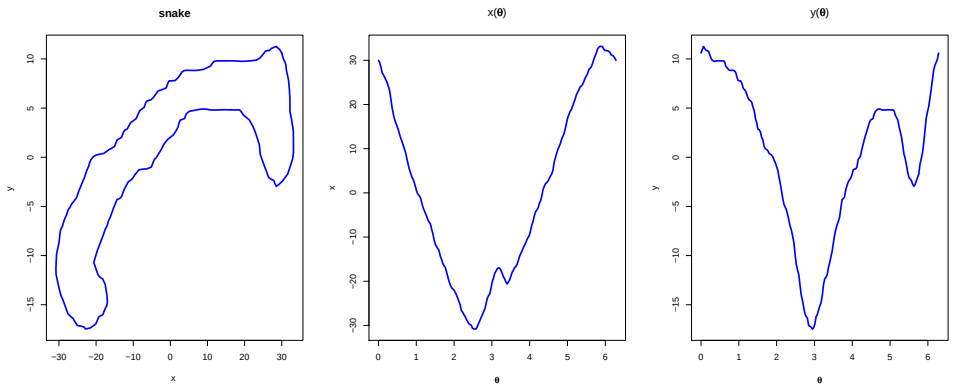

Figure 36: The closed curve on the left (the GVF snake) is decomposed into two functions $x(\theta)$ and $y(\theta)$ (middle and right).

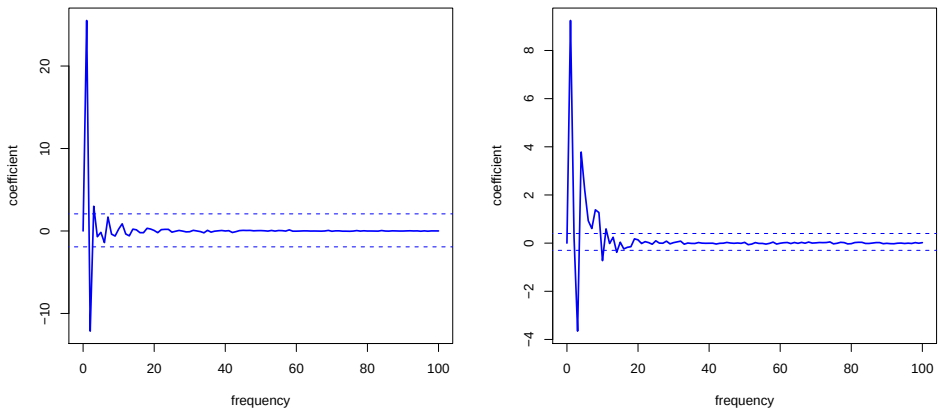

Figure 37: The results of FFT of function $x(\theta)$ (left) and $y(\theta)$ (right) in Figure 36. The thresholds of fast weighted Fourier analysis are given as dashed lines.

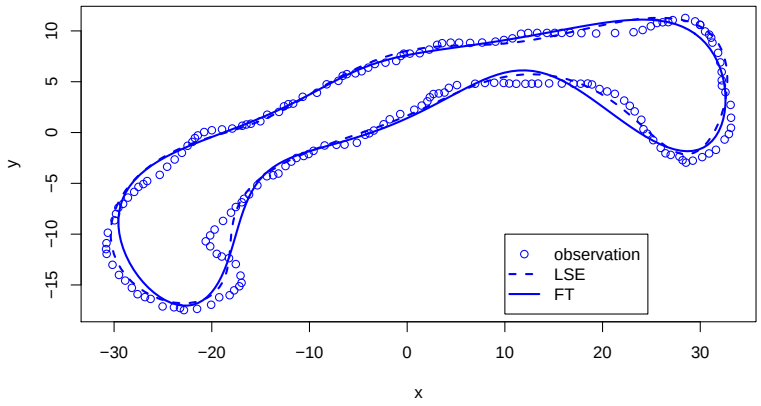

Figure 38: Reconstruction of the snake in Figure 36 using LSE and fast weighed Fourier analysis.

curve $\{p_i\}_{i=1}^n$, we have

$$C(s_i) = (x(s_i), y(s_i)), \quad 0 = s_1 < s_2 \cdots < s_n = 2\pi.$$

Then, we are going to use the fast weighted Fourier analysis on two curve $x(s)$ and $y(s)$, $s \in [0, 2\pi]$. The estimated functions are shown in Figure 36. Figure 37 and Figure 38 show the results of fast weighted Fourier analysis give comparable results to that of LSEs, while fast weighted Fourier analysis using fewer basis functions.

4.5 Two-dimensional fast weighted Fourier analysis

The Fourier transform on the 2-sphere is equivalent to SPHARM (Healy et al., 2003). By (38), one can compute the coefficients of SPHARM using

$$\begin{aligned} \beta_{l,|m|} &= \frac{1}{2}(f_{l,|m|} + f_{l,-|m|}), \\ \beta_{l,-|m|} &= \frac{1}{2}(f_{l,|m|} - f_{l,-|m|}) \end{aligned}$$

for $-l \leq m \leq l$.

4.5.1 Model estimation comparison

The computation time is related to both the number of the observation N and the number of basis functions K as shown in Chapter 2. We first compare the computation time of fast weighted Fourier analysis, LSE and AIR. We study the linear model

$$\boldsymbol{f} = \boldsymbol{Y}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I}).$$

where $\boldsymbol{Y}$ is the $N \times K$ design matrix whose columns are SPHARM basis functions. We first compare the computation time of the three methods to estimate the coefficients using first K basis functions. For this simulation, K ranges from 100 to 400. The comparison results are shown in Figure 39. As we predict, AIR uses less CPU time than LSE. But fast weighted Fourier analysis absolutely needs less CPU time than the other two methods. We then compare the ac-

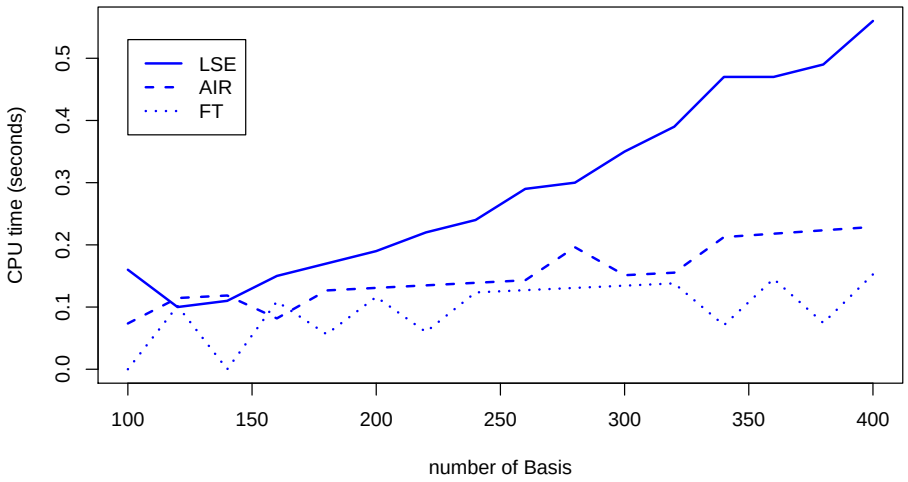

Figure 39: Comparison of CPU times of LSE, AIR and FT.

curacy of the estimations by the three methods. We set up our true model as

$$\boldsymbol{f} = \sum_{i=1}^{I} b_i \phi_{j_i} + \boldsymbol{\epsilon} \tag{43}$$

where $\{\phi_{j_i}\}_{i=1}^{I}$ are selected basis function from the Fourier basis functions. Coefficients b_i's are pre-specified numbers. Normal errors are added to the true model to simulate the observations. Then we estimate the true model based on the observations using the three methods. To characterize the deviation of the estimated model from the true model, we use a L^2-norm, $\hat{f} - \boldsymbol{f}$, where $\hat{f}$ is the estimated model. We repeat the simulation 100 times. The box-plots of the residual sum of squares are shown in Figure 40. We find that the accuracy

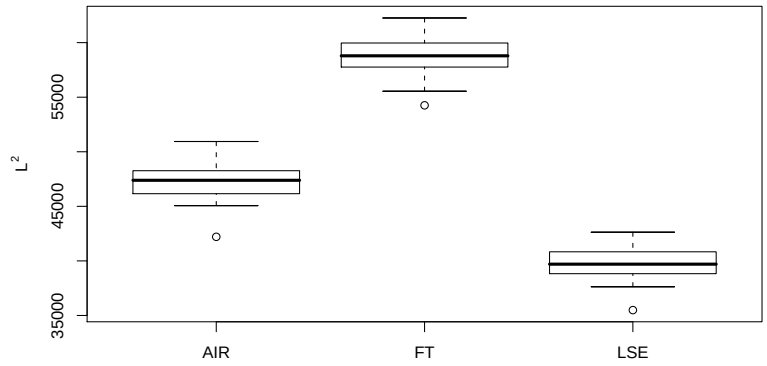

Figure 40: The box-plot of L^2 distances of the simulation that compares accuracy of LSE, AIR and fast weighted Fourier analysis.

of fast weighted Fourier analysis is not as good as LSE and AIR, which is the tradeoff of fast computation. We finally apply the fast weighted Fourier analysis to the mandible surface estimations. We compare the results of fast weighted Fourier series analysis with LSE results. For fast weighted Fourier analysis, we use an average 165×3 (for x, y, z coordinates) basis functions, while LSE uses an average of 324×3 basis functions. We also compared the plots of mandible surfaces obtained from the two methods as shown in Figure 41. The results of the two methods are very close. But fast weighted Fourier analysis uses only about half the basis functions of those for LSE.

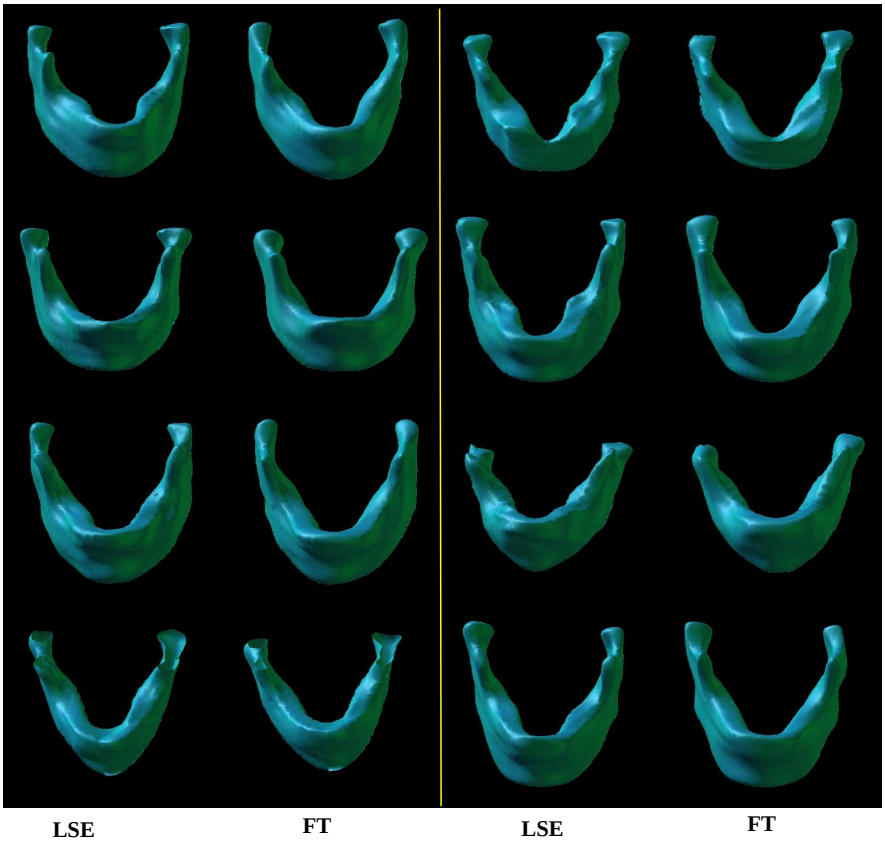

Figure 41: Comparison of Mandible surfaces from LSE and fast weighted Fourier series analysis (indicated by "FT").

4.5.2 Model selection comparison

We also compare the fast weighted Fourier analysis with other model selection methods. In our comparison procedure, we found that some model selection methods, such as AIC and BIC, are extremely slow with the large number of basis functions. Therefore, we only compare our method with two model selection methods that worked reasonably well: LASSO and Dantzig selector method. There are tantalizing similarities between DS and LASSO but they produce different models. Some interesting discussions of the comparison between the two methods can be found in Bickel (2007); Efron et al. (2007).The definition of LASSO can be expressed as an optimization method of finding coefficient β

$$\min_{\beta} \|(y - X\beta)\|_2 \quad \text{subject to } \|\beta\|_1 \leq s,$$

where $\|\cdot\|_2$ is the l^2-norm, $\|\cdot\|_1$ is the l^1-norm, y is the responses, s is a pre-specified threshold and X is the predictor. The definition of the Dantzig selector (DS) can be expressed as

$$\min_{\beta} \|X(y - X\beta)\|_\infty \quad \text{subject to } \|\beta\|_1 \leq s,$$

where $\|\cdot\|_\infty$ is the l^∞-norm. With a bound on the l_1-norm, LASSO minimizes the mean squared error while DS minimizes the maximum component of the gradient of the squared error function. If the threshold s is large so that the constraint has no effect. These two methods produce the identical solution. However, for other values of s, they are somehow different.

	σ	FWFA			LASSO			Dantzig selector		
		AS	AN	T	AS	AN	T	AS	AN	T
5	0.05	4.89	8.05	1.06	5	20.56	47	5	19.26	49
	0.5	4.69	6.74	1.11	5	21.29	49	5	18.21	49
	1	4.64	6.23	0.81	5	19.58	48	5	21.61	48
	5	4.52	6.04	0.77	4.91	18.08	49	4.93	20.12	48
	15	3.03	5.32	0.76	3.55	15.76	49	3.51	14.21	49
90	0.05	90	107	1.77	90	160	274	90	161	172
	0.5	89.23	102	1.80	90	162	276	90	162	172
	1	89.23	102	1.80	90	162	280	90	163	174
	5	88.75	101	1.80	89.77	161	277	89	161	172
	15	77.69	92	1.76	79.11	139	274	79	140	170

Table 3: The model selection comparison of fast weighted Fourier analysis, LASSO and Dantzig selector. 'FWFA' stands for fast weighted Fourier analysis, 'AS' stands for average score, 'AN' stands for average number of predictors selected, and 'T' stands for computation time.

The three methods, fast weighted Fourier analysis, LASSO and Dantzig selector are compared via two simulation studies. In the first simulation, we assume the true model as

$$\boldsymbol{Y} = a_1\phi_{10} + a_2\phi_{30} + a_3\phi_{50} + a_4\phi_{70} + a_5\phi_{90}.$$

Then the observation $\boldsymbol{y} = \boldsymbol{Y} + \sigma * \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, I)$. $\{a_i\}_{i=1}^5$ are pre-specified numbers. We are going to select the true model from the first 100 basis functions $\{\phi_j\}_{j=1}^1 00$. To test the robustness and accuracy of our method against various errors, We use five different σ's (from 0.05 to 15). For every given σ, the three model selection methods are applied to estimate the true model. We repeat model selection procedure 100 times for every σ.

In image analysis, the shapes of the observations are always complex. Therefore, it requires more Fourier basis functions to give a good representation. In the second simulation, the true model has more terms, i.e. I is large in (43). For $I = 90$, we are going to select from the first 225 ($=15^2$) basis functions. We repeat the simulation 100 times.

The results of comparison are shown in Table 3, We see that LASSO and Dantzig selector are very conservative, but only achieve a little better average scores. LASSO and Dantzig selector are much slower than fast weighted Fourier analysis in model selection. Clearly, for the model selections in weighted Fourier analysis, fast weighted Fourier analysis clearly outperforms LASSO and Dantzig selector methods.

Chapter 5

Medical Imaging Applications of Weighted Fourier Series

In this chapter, we are going to apply weighted Fourier series to medical image analysis using the techniques we introduced in previous chapters. We first explore the possibility of developing an automated diagnostic tool for detecting autism based on MRI measurements. We then develop a systematic framework of detecting the regions on amygdala surface where the statistically significant difference in autism is located. A fast weighted Fourier analysis of the growth patterns for mandible surfaces is also proceeded.

5.1 Automated diagnosis of autism

The underlying neuropathology of autism appears to be complicated and undetermined. Various literatures suggested that the abnormalities of the corpus callosum are involved (Piven et al., 1997; Hardan et al., 2000; Chung et al., 2004; Waiter et al., 2005; Alexander et al., 2007). In this section, we are going to develop a regression tree based classification method for automated diagnosis

of autism using weighted Fourier series as a shape descriptor (Golland et al., 1999).

5.1.1 Segmentation

With medical images playing an increasingly important role in the diagnosis and treatment of diseases, the medical image analysis community has become preoccupied with the challenge of extracting useful information about anatomic structures from medical images, since almost all the interesting biomarkers have to be derived from the image segmentation. Segmenting structures from medical images is in general difficult due to the sheer size of the image data sets and the complexity and variability of the images themselves. Deformable models (Kass et al., 1987; Terzopoulos and Fleischer, 1988; McInerney and Terzopoulos, 1996) provide promising and vigorously model-based approach to computer-assisted medical image segmentations. It is widely recognized that the potency of deformable models stems from their ability to segment, match, and track anatomic structures of images by exploiting constraints derived from the image data together with a *priori* knowledge about the location, size, and shape of these structures. Deformable models have been applied to edge detection (Kass et al., 1987), segmentation (Terzopoulos and Fleischer, 1988; Xu and Prince, 1997), motion tracking (Leymarie and Levine, 1993), and nonlinear registration (Davatzikos, 1996; Gefen et al., 2003). We are particularly interested in one dimensional deformable models, the snakes, or the active contours (Kass et al.,

1987).

A snake (Kass et al., 1987; Terzopoulos and Fleischer, 1988) is a deformable curve

$$C(s) = (x(s), y(s)) \in \mathbb{R}^2, \quad s \in [0, 1],$$

which moves within the image and converges to the desired boundary by minimizing the energy functional

$$\begin{aligned} E &= \int_0^1 \frac{1}{2}(\alpha\|C'(s)\|^2 + \beta\|C''(s)\|^2)ds + \int_0^1 E_{\text{image}}(C(s))ds \\ &= E_{\text{int}} + E_{\text{ext}}, \end{aligned}$$

where α and β are the weighting parameters that control the snake's tension and rigidity. The energy functional is divided into two parts: the internal energy E_{int}, which is generated from interaction of the snakes itself and control the smoothness of the snake; and the external energy E_{ext}, which is derived from the images:

$$E_{\text{ext}} = \int_0^1 \|\nabla G_\sigma * I(x(s), y(s))\|^2 ds$$

where ∇ is the gradient operator, and $G_\sigma * I$ denotes the image convolved with a Gaussian smoothing filter whose bandwidth is σ.

To numerically implement the snakes, one usually tries to solve the equivalent Euler equation iteratively

$$\begin{cases} C_t(s,t) &= \alpha C''(s,t) - \beta C^{(4)}(s,t) - \nabla E_{\text{image}}, \\ C(s,t) &= C_0(s) \end{cases}, \tag{44}$$

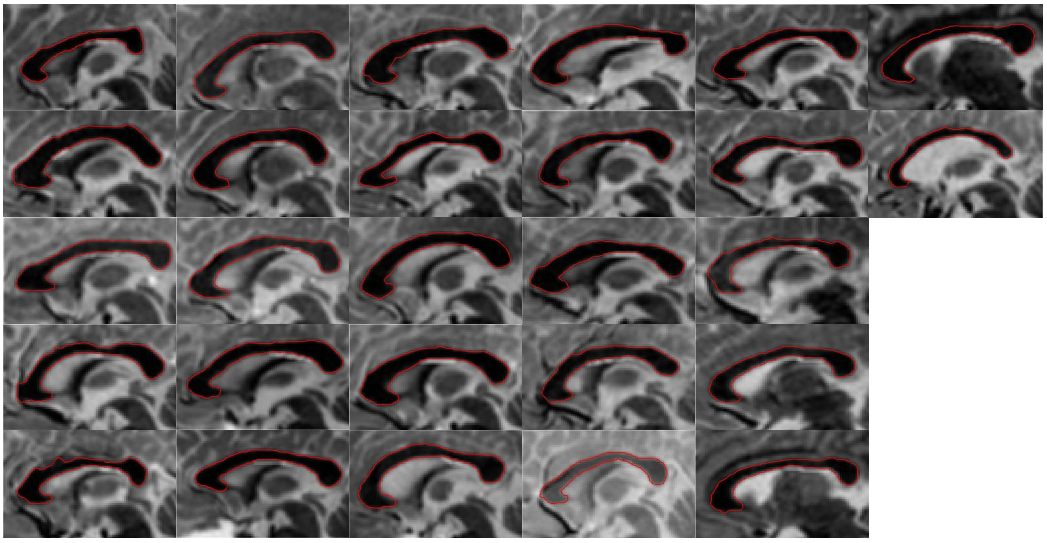

Figure 42: All the 27 GVF snake segmentation results (the red curves) of the corpus callosum data. The background images are cut from the original images for better illustration.

where we call ∇E_{image} the external force. *Gradient vector flow* (GVF) snakes (Xu and Prince, 1997) introduce a new external force f, which minimizes

$$\mathcal{E} = \int\int \mu(\|\nabla v_1\|^2 + \|\nabla v_2\|^2) + \|\nabla f\|^2 \|v - \nabla f\|^2 dxdy$$

where $v = [v_1(x, y), v_2(x, y)]^\tau = \nabla E_{\text{image}}$. GVF snakes distinguish from traditional snakes by being able to converge to the concave parts of the boundaries and capture the detailed information of boundaries (as shown in Figure 42). But the tradeoff of capturing the detailed information of the corpus callosum boundaries is that the snakes are in general noisy. Therefore, a better shape descriptor of the snakes is needed.

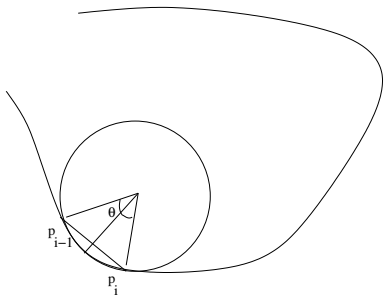

Figure 43: The plot shows the difference of the estimation of arc-length of a curve using curvature-based method and the method using the distance between two points.

5.1.2 WFS representation of the snakes

The curvature calculation using (18) is independent of any parametrization. Thus we are able to use the curvature information to improve the arc-length parametrization procedure, especially when the data is sparse. In Figure 43, let $k(p_i)$ be the curvature at p_i. Since the radius of the circle going through p_i and p_{i-1} is $1/k(p_i)$, by definition,

$$\theta = 2\arcsin(\frac{k(p_i)\|p_i - p_{i-1}\|}{2}).$$

Therefore, the arc-length between p_i and p_{i-1} is $[1/k(p_i)]\theta$. Clearly

$$\begin{aligned} \|p_i - p_{i-1}\| &< \frac{1}{k(p_i)}\theta \\ &= \frac{2}{k(p_i)} \cdot \arcsin(\frac{k(p_i)\|p_i - p_{i-1}\|}{2}). \end{aligned}$$

Therefore, the arc-length parametrization defined in (17) underestimates the true parameters. By using the curvature information, we design an arc-length

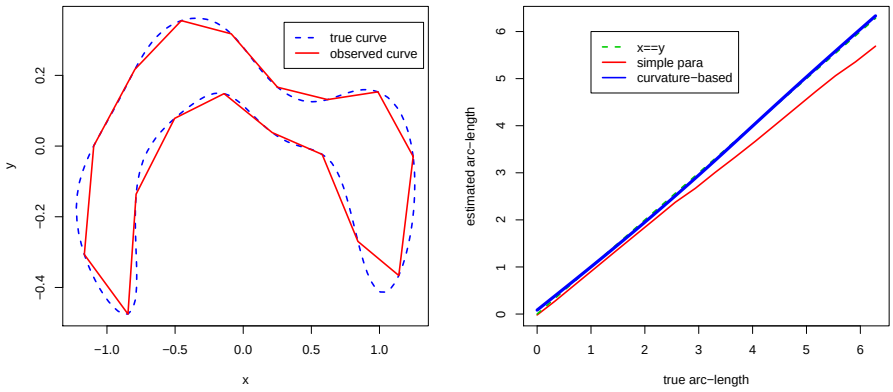

Figure 44: Left, simulated CC boundaries; Right, the comparison of two parametrization results versus true parametrization where the "simple para" stands for the simple parametrization procedure by simply adding the distances between points.

parametrization method as

$$s_i \quad = \quad s_i + \frac{2}{k(p_i)} \cdot \arcsin(\frac{k(p_i)\|p_i - p_{i-1}\|}{2}), \quad i = 1, 2, \cdots, n,$$

where $s_0 = 0$. This method approximates the length of the curve between p_i and p_{i-1} using the arc-length between the two points. The method defined in (17) calculates the length of the straight line between p_i and p_{i-1}. Clearly our method gives a better parametrization.

The curvature computation using the first and the second derivatives is not applicable here since computing the first and second derivatives requires a pre-specified parametrization. This curvature-based parametrization gives a more accurate estimation of arc-lengths than the classic method since it uses

higher order information of the curves (it is equivalent to second order Taylor expansion of the plane curves (Wang, 2003)). This parametrization has an order of convergence $o(h^2)$, while the simple classic parametrization method defined in (17) only has an order of convergence $o(h)$, where h is defined as the maximum of the distances between two neighboring points. Figure 44 shows that the simple parametrization underestimates the arc-lengths, and our method gives a better parametrization (closer to the ground truth).

In practice, the GVF snakes result in noisy and irregularly-spaced closed curves. For example, GVF snakes (Xu and Prince, 1997) allow elastic evolution of curves, which makes the obtained snakes irregularly-spaced. To capture the detailed information of the boundaries of the objects, the snakes become noisy when trying to fit the uneven boundaries. From Figure 19, we know that the curvature functions from noisy and irregularly-spaced curves are also noisy. So it is natural to find their smooth representations of closed curves using WFS. Other smoothing methods might not be applicable. For example, smoothing splines, or local polynomial regression give smooth representations, but these representations are not necessarily periodic (the curvature functions of closed curves are periodic).

In Chapter 2, we introduced an F-statistic based method to choose the proper degrees of WFS representations. For small-sized curve data, a more sophisticated method can be applied. *Discrepancy principle* (DP) method is widely used in the field of experimental medicine for the studies of the glucose

regulation (Morosov, 1966; Eaton et al., 1980; Morosov, 1984; De Nicolao et al., 1997; Hovorka et al., 1998; Sparacino et al., 2001; Toffolo et al., 2001). In those studies, DP was used to choose the optimal tuning parameters of the regularized deconvolution algorithms. For the curve fitting problem, DP chooses the fitted curve such that the discrepancy of the fitting is just equal to the average measurement error. Let the WFS representation of a curvature function $k(s)$ of degree L be

$$\begin{aligned} k_L(s) &= \langle \frac{1}{2}, k(s) \rangle + \sum_{l=1}^{L} e^{-l^2 t} \langle \cos(ls), k(s) \rangle \cos(ls) \\ &\quad + \sum_{l=1}^{L} e^{-l^2 t} \langle \sin(ls), k(s) \rangle \sin(ls). \end{aligned}$$

Under the assumption that the estimated errors are normally distributed, DP chooses L such that

$$(\boldsymbol{k} - \boldsymbol{k}_L)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{k} - \boldsymbol{k}_L) = N$$

where $\boldsymbol{k}$ and $\boldsymbol{k}_L$ are discrete k and k_L, $\boldsymbol{\Sigma}$ is the cross subjects sample covariance and N is the number of observations.

The WFS representations of the noisy and irregularly-spaced hypotrochoids in Figure 19 are calculated with degrees chosen by DP. The estimated smooth curvature functions are shown in Figure 45. From the first three plots, DP gives a very good approximation of true curvature functions. In the forth plots, DP gives a slightly over-smoothed approximation. Overally speaking, DP gives satisfactory results for the curvature approximation. Therefore, for every GVF

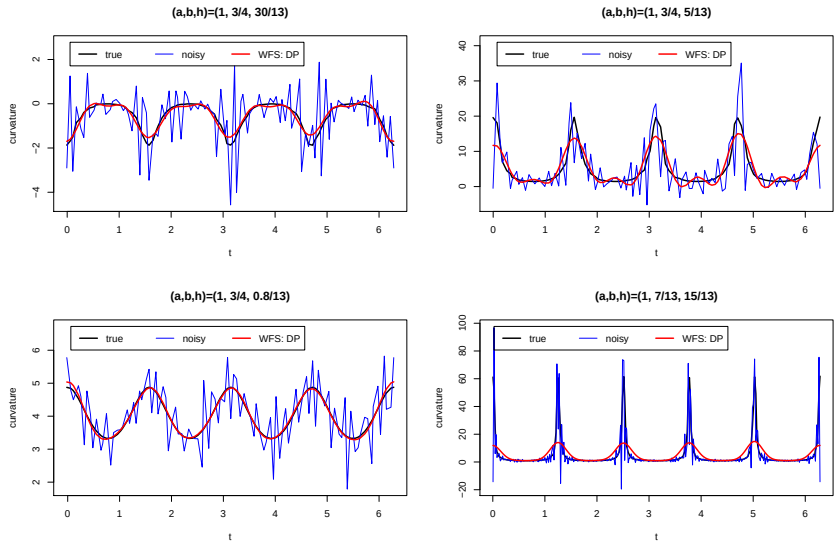

Figure 45: The plots of the WFS representations of the curvature functions that are calculated using DP. The hypotrochoids in Figure 19 are used.

snake (the obtained boundary of a corpus callosum), one first computes the curvatures of the curves (the curvature is usually noisy). Then the WFS representation of its curvature function is computed using DP as shown in Figure 46.

5.1.3 Classification using decision trees

From Figure 42, we see that the snakes are different in locations, sizes and orientations. The snakes are also noisy. In Chapter 2, it is shown that weighted Fourier series is a good shape descriptor. A curvature-based method aligns all the snakes nicely. After the alignment, every snake is represented by a weighted

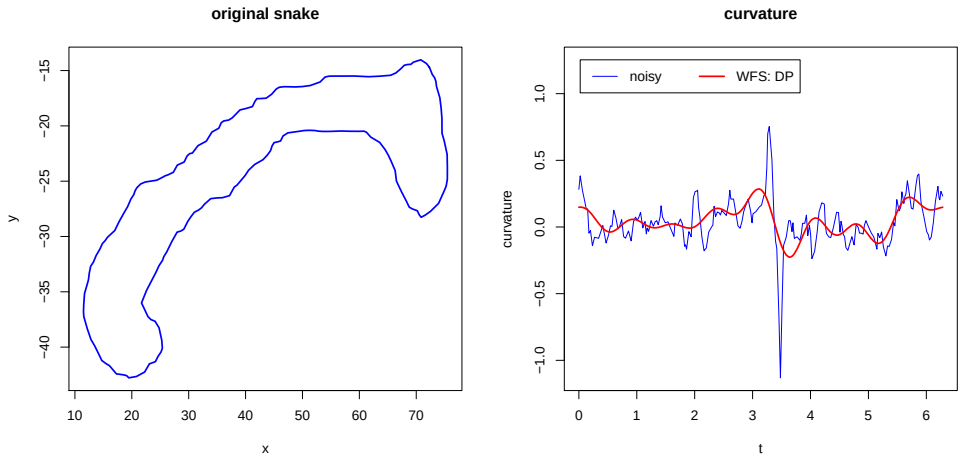

Figure 46: An example of the extracted GVF snake and its corresponding curvature functions.

Fourier series of their curvatures. Therefore, the coefficients of the weighted Fourier series give a multivariate representation of the original snakes.

Decision trees (Breiman et al., 1984) contain a binary question about certain features at each node in the tree. The leaves of the tree contain the best prediction based on a training data. The basic algorithm is given a set of samples to find the best "splits" that minimize certain cost function. The interpretation of the results summarized in a tree is straightforward. Tree methods are nonparametric and nonlinear. Therefore, there are very few assumptions about the data. Another advantage of decision tree methods is that they are usually very flexible on the boundaries. For example, Figure 47 shows why the decision trees are better than *linear discriminant analysis* (LDA).

Decision-tree-based classification techniques (Loh and Shih, 1997; Loh, 2002;

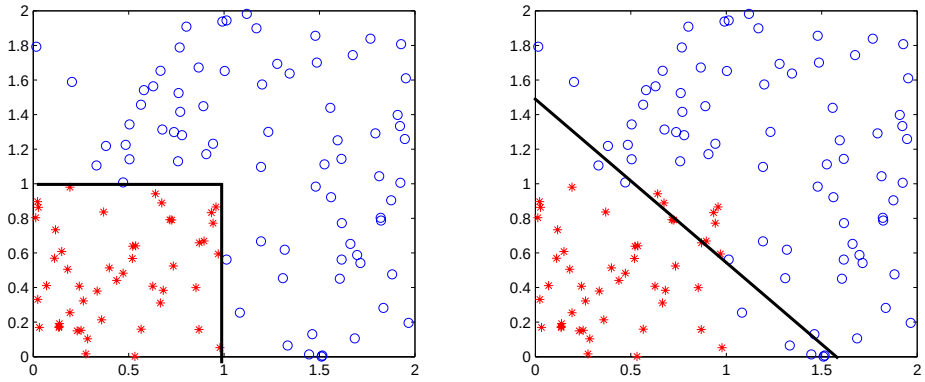

Figure 47: Left: the classification result using a decision tree algorithm; right: the classification result using LDA. The solid lines are the boundaries of two classes. The plots show that decision trees are more flexible on the boundaries than LDA.

Kim and Loh, 2001) were applied to determine if it is possible to differentiate autism purely based on the shapes of CC curves. The following decision tree packages are used: *Classification Rule with Unbiased Interaction Selection and Estimation* (CRUISE) (Kim and Loh, 2001); *Generalized, Unbiased, Interaction Detection and Estimation* (GUIDE) (Loh and Shih, 1997); *Quick, Unbiased and Efficient Statistical Tree* (QUEST) (Loh, 2002). LDA is also used for method comparison.

CRUISE implemented two univariate split methods and one linear combination split method to construct the classification trees with multi-way splits. It is a much-improved descendant of an older algorithm called FACT (Loh and Vanichsetaku, 1988). GUIDE was specifically designed to eliminate variable

Methods	LDA	CRUISE	GUIDE	QUEST
Misclassification rate	0.25	0.22	0.15	0.37

Table 4: The automated autism diagnosis results using LDA and decision tree methods: CRUISE, GUIDE and QUEST.

selection bias, which can undermine the reliability of inferences from a tree structure. GUIDE controls bias by employing chi-square analysis of residuals and bootstrap calibration of significance probabilities. In this way, GUIDE allows fast computation, natural extension to data sets with categorical variables, and automated detection of local interactions between variables. QUEST was designed to overcome the problem with classification trees based on exhaustive search algorithms, which tend to be biased towards selecting variables that afford more splits. Each decision tree algorithm has its strength and weakness. From the study of autism, we find that GUIDE gives the best classification results.

30 different combinations of training sets and test sets are used in this experiment. As shown in Table 4, with a small sample size of 27 subjects, we still manage to achieve an impressive 15% average misclassification rate (85% average correct diagnostic rate). The results are consistent with those of two previous structural imaging studies of autism in corpus callosum (Chung et al., 2004) and (Alexander et al., 2007). With the additional social and behavioral measurements, the correct diagnostic rate might be further improved.

5.2 Autism detection in amygdala

In this section, we show a general procedure of detecting autism using weighted Fourier analysis of amygdala images based on the procedure described in Chung (2006a) and Chung et al. (2006b, 2008a).

5.2.1 Parametrization

High resolution *magnetic resonance images* (MRI) were obtained using a 3-Telsa scanner with a quadrature head coil at the Waisman Laboratory for Brain Imaging and Behavior at the University of Wisconsin, Madison. The details on image acquisition parameters are given in Nacewicz et al. (2006); Chung et al. (2008b). MRIs are reoriented to pathological plane (Convit et al., 1999) for optimal comparison with anatomical atlases. Manual segmentation was done by an expert and the reliability of the manual segmentation was validated by two raters on 10 amygdalae resulting in intraclass correlation of 0.95. Nacewicz et al. (2006) evaluated amygdala volume in individuals with autism spectrum disorders and its relationship to laboratory measures of social behavior to examine the variation in amygdala related to the autism symptom severity. The original segmentation results were saved in the binary format. We first apply Marching Cubes method (Lorensen and Cline, 1987; Styner et al., 2006) to extract amygdala surfaces and their triangulations as shown in Figure 48.

As shown in Chapter 2, a good parametrization of a surface is crucial to the estimation of WFS using iterative regression methods, such IRF and AIR. A

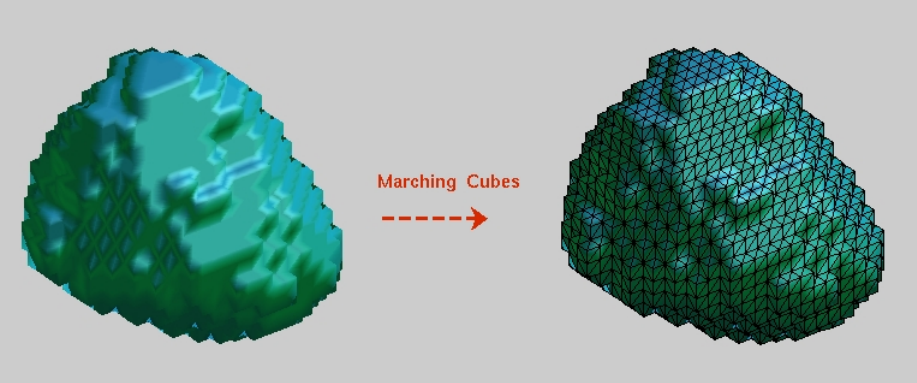

Figure 48: The results of Marching Cubes amygdala boundary extraction.

parametrization of a surface can be viewed as a one-to-one mapping from the surface to certain domain, for example, a unit sphere. Parameterizations have many applications in sciences and engineering, including scattered data fitting (Eck et al., 1995), re-parametrization of spline surfaces (Duren and Hengartner, 1997), and texture mapping (Levy and Mallet, 1998; Zigelman et al., 2001). The most important, parametrization is the foundation for mathematical modeling of surfaces (Brechbuehler et al., 1995; Lee et al., 2002; Gotsman et al., 2003; Styner et al., 2006). After a proper surface parametrization procedure, the amygdala surfaces can be described as L^2 functions on the unit sphere, and thus weighted Fourier analysis can be applied.

Parameterizations almost always introduce distortion in either angles or areas and a good parametrization in applications is the one which minimizes these distortions in some sense. The parametrization problem is in general a constrained optimization problem. The optimal parametrization $((\theta^*, \phi^*))$ is

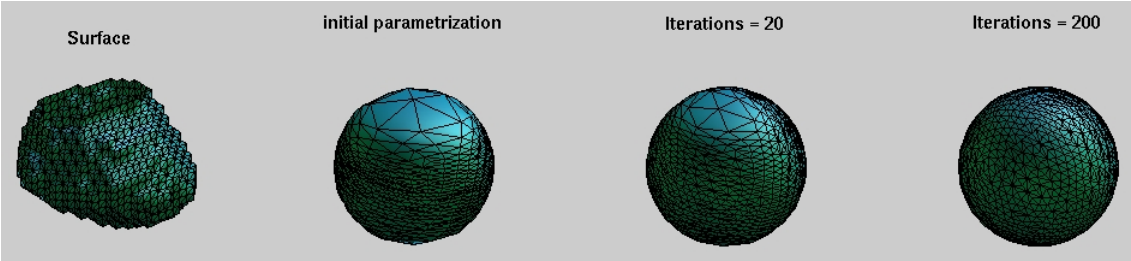

Figure 49: The process of area-preserving parametrization. the first one is a selected amygdala surface. The second surface is the triangular mesh on the unit sphere, which is the initial parametrization that preserves the topology and the connection of the surface.

given by

$$(\theta^*, \phi^*) = \arg\min_{(\theta,\phi)} \mathcal{M}((\theta, \phi)) \quad \text{subject to} \quad \mathcal{V}((\theta, \phi)) \leq 0$$

where $\mathcal{M}$ is the distortion function and $\mathcal{V}$ is the validation function.

We use an area-preserving parametrization proposed in Brechbuehler et al. (1995) and Styner et al. (2006), which maps every triangle to a triangle in parameter space with a proportional area and maps every quadrilateral to spherical quadrilateral (minimal distortion), and keeps the connections and topology of triangulation (validation). For amygdala surface parametrization, we use an area-preserving parametrization package "ShapeTool" (Styner et al., 2006). The iterative area-preserving parametrization results are shown in Figure 49.

Optimal WFS degrees of closed surfaces (as shown in Figure 50) are chosen by DP since the size of the amygdala surface is relatively small. Then all the surfaces are represented by weighted Fourier series. They are properly aligned

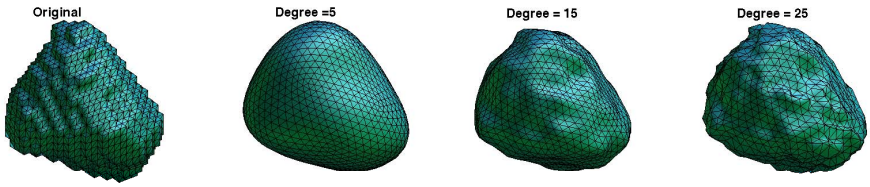

Figure 50: WFS representation of different degrees with t=0.0001. DP choose the optimal degree =15.

by the curvature-based registration method as specified in Chapter 3. The affine aligned amygdalae are shown in Figure 51.

5.2.2 Multiple comparison using random field theory

The studies investigating the development of the corpus callosum in autism have provided mixed results (Alexander et al., 2007). The investigations into the amygdala volumetry are not consistent either (Aylward et al., 1999; Haznedar et al., 2000; Sparks et al., 2002; Nacewicz et al., 2006). Imaging studies (Baron-Cohen et al., 1999; Pierce et al., 2001; Dalton et al., 2005b) have found difference in amygdala activation to faces in individuals with autism. Nacewicz et al. (2006) examined relations between amygdala volumes and quantitative measures of faces processing and gaze fixation. They reported the first relationship between amygdala structure and current and past measures of social impairment in autism. In this section, we are going to detect and localize the shape difference between the autistic and normal amygdala surfaces.

Suppose $\mathcal{S}_1$ and $\mathcal{S}_2$ are the mean surfaces of the autistic amygdala surfaces

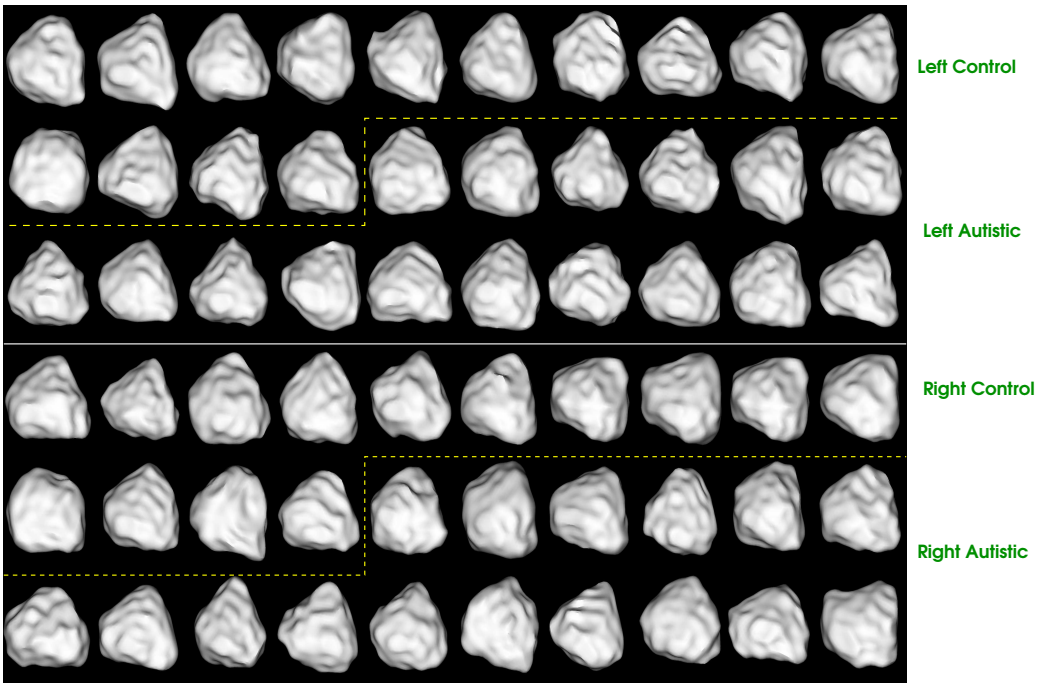

Figure 51: Registered amygdala surface using curvature-based method.

$\{\mathcal{S}_1^j\}_{j=1}^m$ and normal amygdala surfaces $\{\mathcal{S}_2^j\}_{j=1}^n$. To detect the shape difference between autistic amygdalae and controlled amygdalae, we are interested in the following hypothesis test:

$$\begin{aligned} H_0 &: \mathcal{S}_1 = \mathcal{S}_2; \\ H_1 &: \exists p_i \quad \text{such that} \quad \mathcal{S}_1(p_i) \neq \mathcal{S}_2(p_i). \end{aligned}$$

To compare surfaces, we first characterize the random variable of the difference between two multivariate random variables for every point on the surface. In this study, Hotelling's two-sample T^2 statistic is used to model the difference in mean between two multivariate variables (Worsley, 1996; Cao and Worsley, 1999). It is a generalized version of t-statistic. Suppose we have two multivariate samples $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$ and $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n$ in $\mathbb{R}^3$. Hotelling's two-sample T^2 statistic is defined as

$$T^2 = \frac{mn(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}})' \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}})}{m+n}.$$

Hotelling's T^2-statistic is essentially an F-statistic

$$\frac{m+n-p-1}{(m+n-2)p} T^2 \sim F_{p,m+n-p-1},$$

where p is the dimension of the samples.

The maximum $\max T^2$ of all T^2 over a search region (usually it is the entire surface) is used to test for the local differences in mean at an unknown location on the surface (Cao and Worsley, 1999). We want to choose a threshold Z_0 to exclude false positives with high probability (0.95), i.e., a small p-value

$$P(\max T^2 > Z_0) = 0.05.$$

We need to figure out how to compute the distribution of $\max T^2$. Assuming all points are independent and using Bonferroni correction to approximate the distribution of $\max T$, we have

$$P(T^2 > Z_0) \approx \frac{\alpha}{N},$$

where α is the significance level (0.05 for this case) and N is the number of points on the surface. But Bonferroni correction is usually too conservative, especially when the number of points N is large. Most image surfaces are locally correlated. Therefore the assumption of independence can not be applied.

There is no exact result for the null distribution of $\max T^2$ (Cao and Worsley, 1999; Worsley, 2001; Taylor and Worsley, 2007). But for a high threshold Z_0, we can use the random field theory to approximate the probability that $\max T^2$ exceeds Z_0 using expected *Euler Characteristic* (EC) (Worsley, 1996; Cao and Worsley, 1999). The expected EC leads directly to the expected number of clusters above the given threshold, which can be used to approximate the p-value $P(T^2 > Z_0)$.

It is important to find an appropriate representation for the EC at every point of the surface. In that way, one writes the EC in locally defined terms of certain random field. This representation comes from Morse theory (Worsley et al., 1995). The expected EC becomes the expectation of the determinant of the second derivatives of the random field. Worsley (1996); Cao and Worsley

(1999) showed that, in probability one, the distribution of $\max T^2$

$$P(\max T^2 > z) \approx \sum_{d=0}^{D} \text{Resels}_d \cdot \text{EC}_d(z),$$

where D is number of maximal dimension in the search region, EC_d is the d-dimensional EC density, and Resels_d is the number of d-dimensional resels. Resel is a measure of the "resolution size" in the statistical map,

$$\text{Resels}_d = \frac{V}{\text{FWHM}^d}.$$

In Chpater 2, we showed how to calculate FWHM of heat kernel numerically. By using the formulas of Hotelling's T^2 field in Cao and Worsley (1999), we have

$$\begin{aligned} P(\max T^2 > t) \approx\ & 2\int_t^\infty (\rho_0(t) + \frac{\text{Area}}{\text{FWHM}^2} \cdot \frac{(4\log 2)^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \\ & \cdot\rho_0(t) \cdot \frac{(n-1)mt - n(m-1)}{m(1+mt)})dt, \end{aligned}$$

where ρ_0 is the density function of $F_{m,n}$-distribution. For bandwidth $t = 0.01$ and FWHM=0.6262, the density function of $\max T^2$ for the amygdala surfaces and the 0.05 significant threshold are shown in Figure 52.

Since we can calculate Hoteling's T^2 at each point, then using the distribution of $\max T^2$, we have the corrected p-value at each point of the surfaces. Figure 53 shows the multiple comparison results based on the distribution as shown in Figure 52. It is very interesting to find out that there is no significant difference in left amygdala between autistic and control groups. There are no estimated T^2 values that are larger than the 0.05 threshold (≈ 8.5). However, there is significant difference on right amygdala between autistic and control

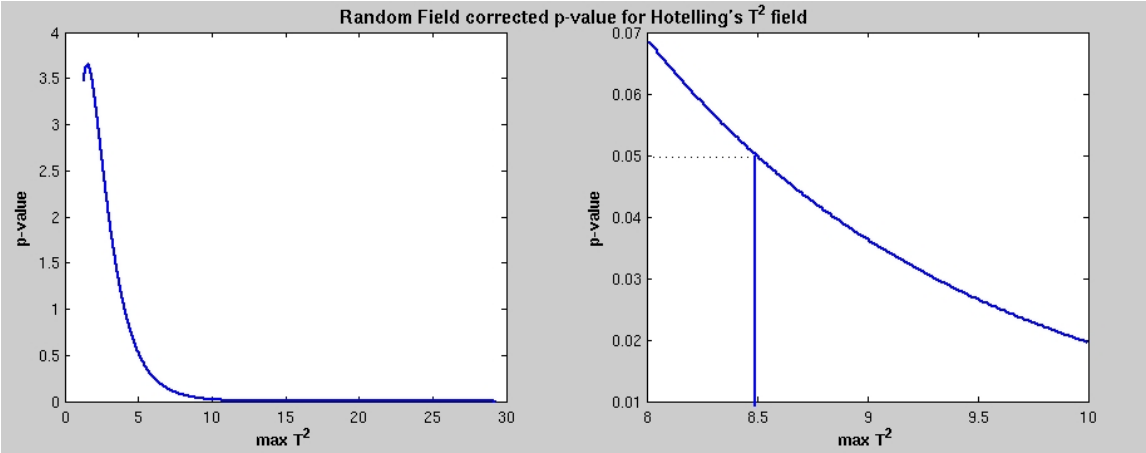

Figure 52: The density function and its 0.05 significant threshold with t=0.01 and WFS degree =15, FWHM =0.6262 and Hotelling's T^2-distribution with degree of freedom (3, 26).

groups since the largest T^2 is larger than 12 and the 0.05 threshold is about 8.5 as shown in Figure 52.

The results are quite interesting that we find significant difference in right amygdala between the normal and autistic groups, which is consistent with the result of a recent research in autism using the same amygdala data in Nacewicz et al. (2006), who found significant difference in individual volumes between autistic group and normal group in right amygdala.

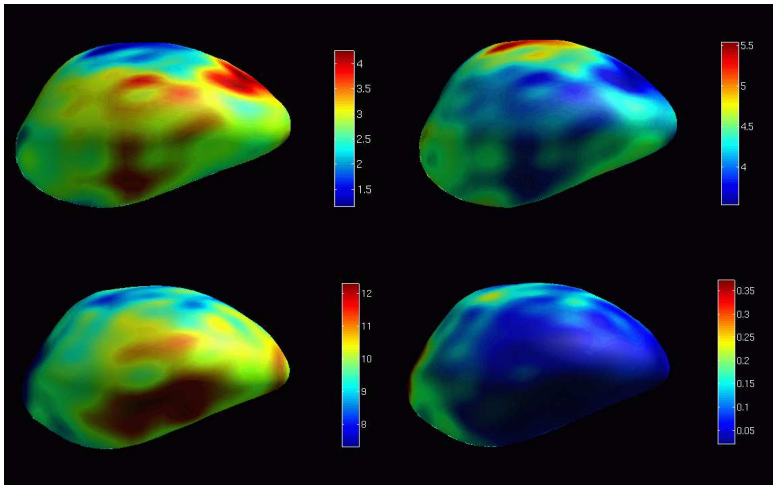

Figure 53: First row: left, the values of Hotelling's T^2 on the mean left amygdala surface; right, the corresponding p-values; second row: left, the values of Hotelling's T^2 on the mean right amygdala surface; right, the corresponding p-values.

5.3 Mandible surface modeling using fast weighted Fourier analysis

The oral and pharyngeal cavities and structures undergo changes in size, shape, and relative proportions during the growth process from infancy through early childhood and adolescence, to adulthood. Acoustic theory indicates that vocal geometry is predictive of the spectrum shape of speech sounds (Vorperian et al., 1999). Various biomarkers from vocal tract region are extracted and measured using MR images (Vorperian et al., 1999). We are especially interested in the growth patterns of the soft tissue and bony vocal tract structures. Growth curves using various models, from piecewise linear model to polynomial fittings

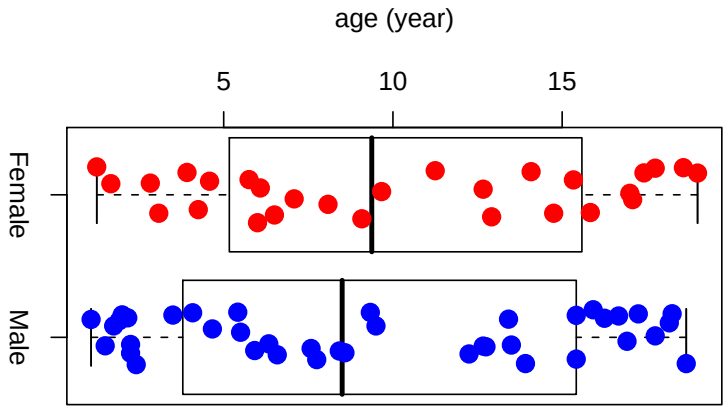

Figure 54: The age distribution of the mandible data. The red points represent female ages and the blue ones represent male ages.

were studied (Vorperian et al., 2005, 2006). A very interesting but challenging problem is modeling the growth pattern of 3D structures, such as mandible surfaces.

In this section, we will study the growth pattern of mandible surfaces using the fast weighted Fourier analysis. 19 female subjects and 33 male subjects are used for this study. The ages of the subject are nicely distributed from 13 months old to 19 years old, which cover the time from an infant to an adult. The distribution of the ages is shown in Figure 54. The mandibles were manually segmented from the original MR images by the researchers from Vocal Tract Development Lab, Waisman Center at the University of Wisconsin at Madison. "ShapeTool" package (Styner et al., 2006) was used to extract the

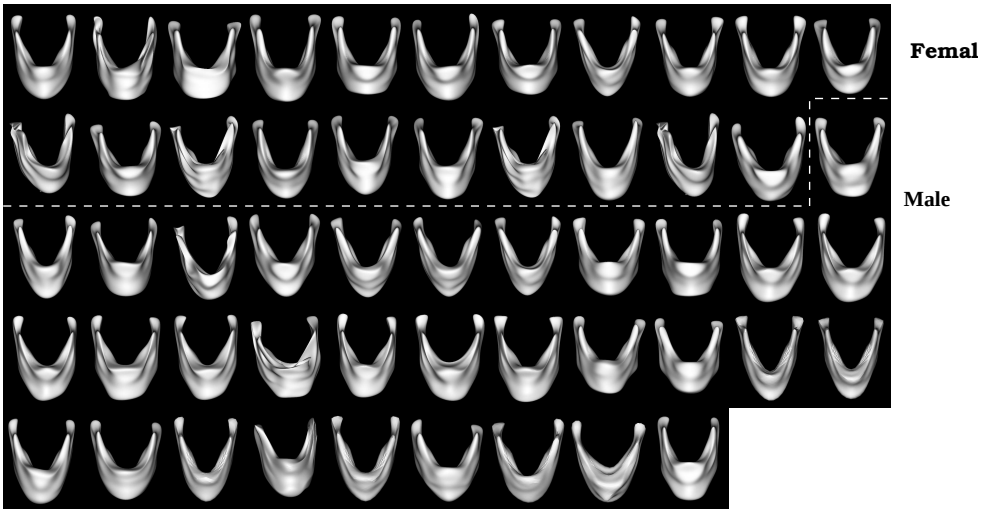

Figure 55: All the registered mandible surfaces. The male and female mandible surfaces are separated by the dashed lines.

mandible surfaces from the segmentation results. Area-preserving parametrization method in Brechbuehler et al. (1995); Styner et al. (2006) is applied. We then use curvature-based registration to align all the mandible surfaces. Since this study will investigate the growth patterns of the mandible surfaces, the sizes of the mandibles are supposed to be different from an infant to an adult. Our model needs to characterize this difference. Therefore, the surfaces are not normalized according to their sizes during the alignment procedure. The registered mandible surfaces are shown in Figure 55.

After registration, we apply fast weighted Fourier analysis method to mandible surfaces to find the WFS representations. The results of fast weighted Fourier analysis are compared with LSE results. For fast weighted Fourier analysis, we use an average of 165×3 (for x, y, z coordinates) basis functions, while LSE uses

an average of 324×3 basis functions. We also compare the plots of mandible surfaces obtained from the two methods as shown in Figure 41. In this figure, we show that the fast weighted Fourier analysis gives comparable results with that using LSE.

Unlike the biomarkers used in Vorperian et al. (2005, 2006), it is not easy to visualize the rough growth pattern and the amount of growth from the scatter plot. We need to define new metrics and new models to represent the amount of growth and the growth patterns.

The registered mandible surfaces are properly aligned and centered. All the mandible surfaces are mapped to a common parameter space. we can define a metric that measures the growth from the mandible surface of the infants for every point in the parameter space. For every point (x, y, z), we define the growth metric as

$$M((x, y, z)) = \sqrt{(x - x_m)^2 + (y - y_m)^2 + (z - z_m)^2}.$$

where (x_m, y_m, z_m) is the coordinate of the corresponding point of 13 months old mandible surface (the youngest we have). We are going to study the pattern of the amount of growth using this metric. We have the ages of all subjects $\{t_i\}_{i=1}^n$, and the metrics of all subject $\{M_i\}_{i=1}^n$. To estimate the underlying growth patterns of the metrics, we fit a smoothing spline f such that f minimizes the penalized residual sum of squares as

$$\sum_{i=1}^{n} (f(t_i) - M_i)^2 + \lambda \int (f''(t))^2 dt, \tag{45}$$

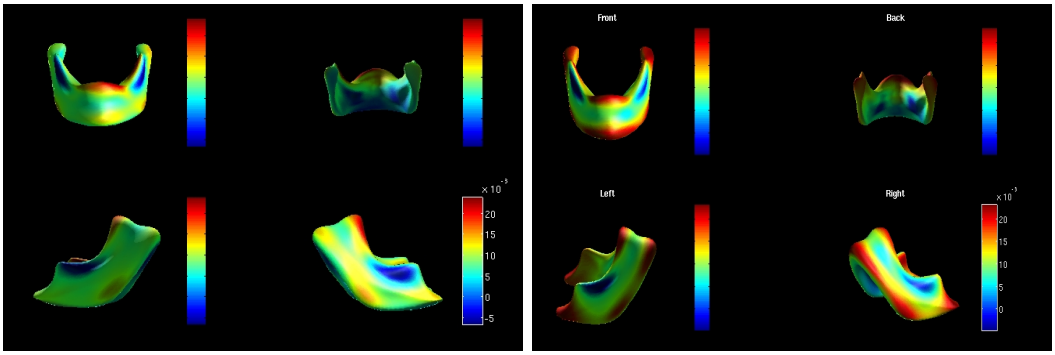

Figure 56: The colormaps of mandible metric growth for females and males. The color indicates the amount of the metric growth. The left plot shows the colormaps of the female mandible metric growth and the right plot shows the colormaps of the male mandible growth. The colormaps are also shown from different view points to give the full information of the metric growth. The units are in millimeters.

where λ is the smoothing parameter that measure the rate of exchange between the fit to the data and the variability of f. The most common computational techniques for smoothing splines is using an order four B-spline (de Boor, 1978) basis function expansion with knots at the sampling points to minimize (45) with respect to the coefficients of the expansion (Chambers and Hastie, 1992; Ramsay and Silverman, 2002). The smoothing spline is estimated by the generalized cross-validation method (Wahba, 1990).

The growth metrics are fitted for every point in the parameter space. This defines a growth metric field that varies smoothly along ages. The colormaps on the mean mandible surfaces show different growth patterns at different parts of the mandibles as shown in Figure 56. We see similar growth patterns at most

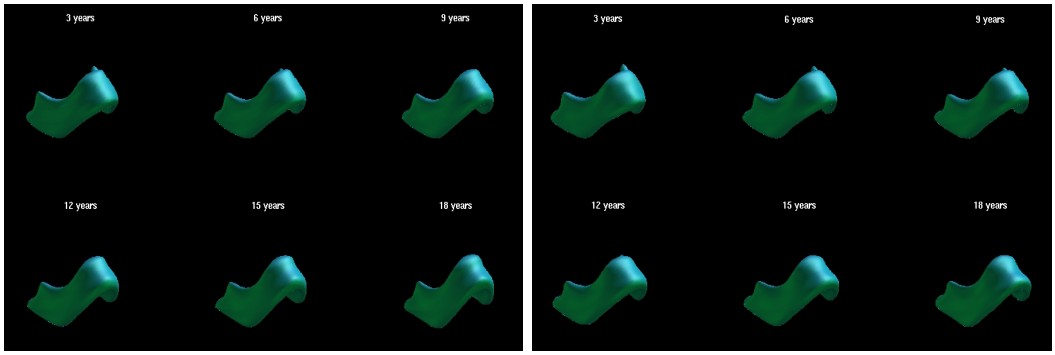

Figure 57: The left plot is the predicted female mandible surfaces and the right plot is the predicted male mandible surfaces. The mandible surfaces are predicted at age 2, 4, 6, 10, 13, and 17 years old.

parts of the mandible surfaces for females and males. From both female and male mandible metric growth colomaps, one can see that rapid growth happens at outer parts of the mandibles and slow growth, or contraction happens at the inner parts of the mandible. The mandible growth also differs between genders. For example, one can find that the front bottom part of male mandibles grows more than the same part of female mandibles does.

We can also characterize the geometric changes of the mandibles. For every point, we have a vector of all x-coordinates, all y-coordinates, and all z-coordinates from all the subjects. Similar to the study of growth pattern of metrics, by using the age information of all the surfaces, we fit cubic smoothing splines to find the growth patterns of x's, y's and z's. From the growth pattern models, we can predict x's, y's and z's at all ages and the shapes of the mandible at all ages, which are shown in Figure 57.

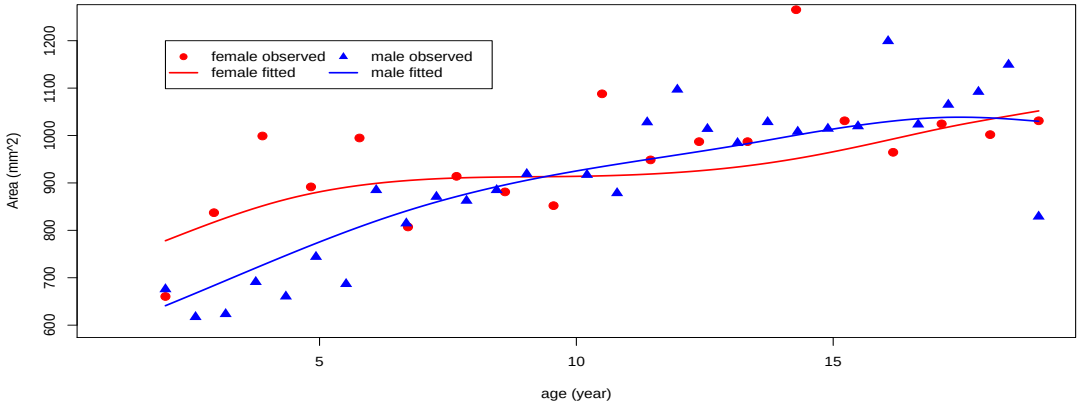

Figure 58: The observed and fitted mandible area growth patterns.

Using the WFS representations (basis functions selected by fast weighted Fourier analysis), one can also characterize the growth curves of mandible surface areas. Surface areas are calculated for every mandible surface based on their WFS representations. The growth curves of female and male mandible surface areas are fitted using cubic smoothing splines as shown in Figure 58. The fitted curves show some interesting facts. By the definitions of neural and somatic growth curves (Scammon, 1930), the growth curve of male mandible areas seems to be a neural growth curve and that of female mandible areas seems to follow a somatic growth curve.

Chapter 6

Conclusions and Discussions

6.1 Summary

In this dissertation, we investigated a systematic framework of medical image analysis using a novel shape descriptor: weighted Fourier series (WFS). WFS is closely related to heat kernel smoothing (Chung et al., 2005; Chung, 2006b). A special case of WFS was formulated as the solution to the heat equation on the unit sphere with given initial conditions (Chung et al., 2006b). We introduced WFS as the unique solution to a more general Cauchy problem, which is based on a non-degenerate self-adjoint linear operator. We provided the theoretical background of WFS and characterized WFS kernel as a classic integral kernel. By Ascolli-Arzela theorem, WFS is also the solution to a special case of Sturm-Liouville problem with initial conditions.

We validated WFS by various simulations. WFS techniques were also applied to the study of autism for automated diagnosis and detection of autistic regions. WFS was also applied to mandible surface modeling. We concluded that WFS has the following properties and advantages (over Fourier series)

- WFS is both a fitting procedure and a smoothing procedure. Fourier series is a special case of WFS. Therefore, WFS is more flexible than Fourier series and can be adjusted according to various situations;
- WFS reduces the Gibbs phenomenon in Fourier series approximation by adjusting the bandwidth;
- WFS is robust for the normality assumption in its related linear models;
- It is relatively easy to compute the smoothness of the WFS kernel in the random field theory (Worsley, 1996; Cao and Worsley, 1999).

Even though the theoretical framework of WFS is well-established, the numerical implementation and computation of WFS can be troublesome for large data, where one has to solve a large linear system. LSE provides an optimal, unbiased and robust estimator for general linear systems. But we showed that LSE is computationally inefficient for solving large linear systems. A stepwise regression algorithm, IRF decomposes a large linear system to a set of small linear systems. IRF then estimates the coefficient of WFS iteratively. It is in general very fast. But IRF does not consider the linear dependency between the small linear systems, which causes inaccurate estimations. We proposed an adaptive stepwise regression method, AIR, which is based on an extra correction step of IRF by reducing the linear dependency of the small linear systems. AIR's computational efficiency is comparable with IRF. But it provides more robust and accurate results.

In previous Fourier series literature (Gerig et al., 2001, 2002; Bulow, 2004; Gu et al., 2004; Shen and Chung, 2006), the optimal degree selection has not been addressed. The degrees were simply selected based on a pre-specified error bound that depends on the size of anatomical structure. For the purpose of finding a stopping rule and model selection, we proposed a method to select the degrees of WFS based on an F-statistic, which uses AIR estimation. We proved that this method is more accurate than the method using an F-statistics based on IRF estimation. We also found that this method improves the power of the underlying hypothesis tests for model selection methods based on stepwise regressions.

Registration plays a key role in medical image analysis. It is a necessary step to remove the translation and orientation difference between images before any comparison and modeling of images could be correctly made. By the fundamental theorem and Bonnet's existence and uniqueness theorem (Stoker, 1969; doCarmo, 1976; Hsiung, 1981; Rubin, 1991), curvature information is independent of locations and rotations and gives a unique representation of a plane curve or a surface. More importantly, this representation is given in the form of lower dimensions than the coordinate representations. This property is crucial to medical image analysis that usually deals with large-sized image data. This enables us to design a curvature-based method to make the image registration computationally more efficient. Therefore, curvature functions represent the data in a parsimonious form and makes the image registration computationally

more efficient.

For curve curvature estimation, we proposed a method that purely depends on the local geometric shapes of the curve. Therefore a curve parametrization is not necessary. It allows us to improve the curve parametrization results by using the curvature information. Our simulations showed that our proposed curvature estimation method is superior to the classic method in robustness and accuracy. For curve data, we showed that we can apply a more sophisticated discrepancy principle degree selection method. We then applied a global shift registration method to align all the estimated curvature functions. Since the registration is purely based on the curvature information, it is much more computationally efficient. To further improve the alignment results, we also applied an elastic curve warping method, which potentially can be applied to any other curve or surface non-linear registration.

Using the curvature information to represent the surface reduces the dimensionality of the surface registration. This is even more important comparing with curve registration since surface data are usually large and complex. Using the recurrence properties of the WFS basis, we proposed a robust and fast curvature estimation method, which is analytically derived from the WFS representations of the surfaces. Then a curvature-based surface alignment is proposed. Our simulations showed it provides comparable results with Procrustes alignment but it is computationally more efficient.

We also introduced an alternative tool to the weighted Fourier analysis: the

fast weighted Fourier analysis, which is closely related to weighted Fourier analysis but approaches the problem from a different angle by using fast Fourier transforms (FFT). We first investigated the linear dependency among the Fourier basis functions. Then we designed a model selection procedure that automatically selects the important basis functions for WFS representation. This method requires fast WFS coefficient calculation. We incorporated FFT to our coefficient estimations. We call this procedure the fast weighted Fourier analysis, which is not only a model selection tool, but also a curve and surface modeling tool. Our simulations showed that fast weighted Fourier analysis provides comparable results with those of LASSO and Dantzig selector, but clearly outperforms these two methods in computational efficiency.

Finally, we showed that weighted Fourier analysis can be applied to various medical image studies. We first explored the possibility of developing an automated diagnostic tool for detecting autism based on MRI measurements. We then developed a systematic framework of detecting and localizing the regions on amygdala surface where the statistically significant difference exists. A fast weighted Fourier analysis of growth patterns of mandible surfaces was also proceeded. By using a decision tree based method, with a small sample size of 27 subjects, we still managed to achieve an impressive 15% average misclassification rate (85% average correct diagnostic rate). The result is consistent with the results of two previous structural imaging studies of autism in corpus callosum (Chung et al., 2004; Alexander et al., 2007). With the additional social

and behavioral measurements, the correct diagnostic rate might be improved. The results of automated detection of autism using amygdala data are quite interesting that we found significant difference in right amygdala between the normal and autistic groups. This result is consistent with the result of a recent research in autism using the same amygdala data in Nacewicz et al. (2006), who found that the volumetric difference between the autistic normal groups in right amygdala is larger than that in left amygdala. Nacewicz et al. (2006) also found significant difference in volume in both left and right amygdala, whereas our results only found significant shape different in right amygdala. Using fast Fourier analysis, we can characterize the growth of the mandible surface in various ways. We measured the local growth of mandible surfaces using a pre-specified metric. We also derived the growth process of the mandible surface using cubic smoothing splines. Mandible surface area growth curves were also fitted based on the observed mandible surfaces.

6.2 Discussions and future works

In Chapter 2, an adaptive regression method, AIR was proposed for the estimation of WFS representations. But clearly AIR has the potential to be applied to other large linear systems. AIR carries out an orthogonalization step further so that it is insensitive to the design matrices. Therefore, one can combine AIR with many model selection algorithms, such as AIC, BIC, LASSO, Dantzig selector and so forth. Using the same idea, one can divide a large model selection

problem to a set of small model selection problems. The linear dependency between those small model selection problems can be reduced by an orthogonalization step. Then one first performs the model selection on every small system as a pre-screening procedure (Fan et al., 2008), then a further selection step can be made based on the selected models of all the small systems.

In this section, we focus on the possible future works of weighted Fourier analysis in medical images. We focus on higher dimensional weighted Fourier analysis and curvature-based nonlinear surface registration.

6.2.1 Higher dimensional weighted Fourier analysis

As we mentioned in Chapter 5, the parametrization process is crucial to WFS analysis since:

1. a good parametrization gives a good approximation of the one-to-one mapping between two topologically equivalent manifolds, such as a genus 0 surface and a 2-sphere;

2. the goodness of parametrization results is one of the most important factors of the performance of stepwise regression methods such as IRF and AIR.

Therefore, parametrization is the foundation of WFS analysis of 2D or 3D medical images, where the geometric features are topologically equivalent to S^1 or S^2. Theoretically, the topology of geometric subjects that are equivalent to S^3

is much more complex. The parametrization of such subjects is essentially the famous Poincaré conjecture.

Theorem 6.1. *Every simply connected compact 3-manifold (without boundary) is homeomorphic to a 3-sphere.*

This conjecture was first proposed in Poincaré, 1904 and subsequently generalized to the conjecture that every compact n-manifold is homotopy-equivalent to the n-sphere if and only if it is homeomorphic to the n-sphere. The generalized statement reduces to the original conjecture for $n = 3$ (Weisstein, 2002). This is one of the Clay Mathematics Institute's $1 million prize problems and many mathematicians have been working on this difficult problem for years (Weisstein, 2002; Robinson, 2003; Collins, 2004).

Nevertheless, with all the present difficulty of higher dimensional Fourier analysis, several groups have made effort to generate the idea of Fourier analysis to four-dimensional space. The four-dimensional version of spherical harmonics, hyper-spherical harmonics have long been an analytical and computational tool for an n-body quantum system (Mitchell and Littlejohn, 1997). Matheny and Goldgof (1995) extended the method to surface harmonics defined on domains other than the sphere and to four-dimensional spherical harmonics. These harmonics enable us to represent shapes which cannot be represented as a global function in spherical coordinates, but can be in other coordinate systems. Bonvallet et al. (2007) proposed a novel shape descriptor based on four-dimensional hyper-spherical harmonics. Shape descriptor using hyper-spherical harmonics

presents benefits of being insensitive to noise, orientation, scale and translation.

Therefore, a four-dimensional WFS, or weighted hyper-spherical harmonics can potentially be developed accordingly. In medical image analysis, four-dimensional weighted Fourier series may be applied to volumetric subject modeling based on an appropriate parametrization. It could provide an analytical and smooth representation of 3D volumetric subjects, such as the whole brain. It could also be used for 3D subject registration.

6.2.2 Non-linear curvature-based registration

Affine alignment tries to map the two surfaces globally. Nonlinear registration allows the alignment of data sets that are mismatched in a nonlinear or nonuniform manner. It is natural to use nonlinear registration to deal with misalignment that can be caused by a physical deformation process, or can be due to intrinsic shape differences. But usually, nonlinear registration is theoretically complex and computationally time-consuming. Due to the complexity of surfaces, a global optimization can not be achieved. In general the surfaces are not convex and thus the functionals defined on these surfaces are not convex either. Therefore, affine alignment is a necessary step before non-linear registration to improve the matching. In this section, we propose a non-linear registration method, which optimally maps the two surfaces locally, but is also constrained by its global patterns by penalizing the curvature mappings. The results of the proposed methods are not convergent now. Further investigation

and validation have to be done.

Given a template surface $\mathcal{S}_0$, one tries to register surface $\mathcal{S}_1$ using an optimal transformation $\Phi^* : L^2(S^2) \to L^2(S^2)$, which is the solution to the following functional

$$\arg\min_{\Phi} \int_0^{2\pi} \int_0^{\pi} \|\Phi(\mathcal{S}_1) - \mathcal{S}_0\|^2 \sin\theta d\theta d\phi.$$

Even though this is very intuitive and straight forward. But this transformation could be non-smooth (Beg et al., 2005). The optimal transformation is the one that minimizes the cost function with proper smoothness. Therefore, we propose a curvature-based non-linear registration method. Let $\mathcal{C}(\mathcal{S})$ denotes the curvature field of surface $\mathcal{S}$. The curvature-based registration Φ^* is the solution to the optimization

$$\arg\min_{\Phi} \int_0^{2\pi} \int_0^{\pi} (\|\Phi(\mathcal{S}_1) - \mathcal{S}_0\|^2 + \lambda\|\mathcal{C}(\Phi(\mathcal{S}_1)) - \mathcal{C}(\mathcal{S}_0)\|^2) \sin\theta d\theta d\phi. \quad (46)$$

We implement the registration method in an iterative fashion. Each time, we improve our registration in a small neighborhood of the surfaces

$$\arg\min_{\Phi_\delta} \int_0^{2\pi} \int_0^{\pi} (\|\Phi_\delta(\mathcal{S}_1) - \mathcal{S}_0\|^2 + \lambda\|\mathcal{C}(\Phi_\delta(\mathcal{S}_1)) - \mathcal{C}(\mathcal{S}_0)\|^2) \sin\theta d\theta d\phi$$

where

$$\Phi_\delta(\mathcal{S}_1)(\theta,\phi) = \mathcal{S}_1(\theta',\phi'), \quad (\theta',\phi') \in B_\delta((\theta,\phi))$$

where $B_\delta((\theta,\phi))$ is the ball with center (θ,ϕ) and radius t. A small δ is usually chosen for better numerical implementation. We can show that the transformation Φ defined in (46) is a smooth transformation. First, the functional in (46)

can be divided into two parts

$$\begin{aligned} E_{\text{int}} &= \int_0^{2\pi}\int_0^{\pi} (\|\Phi(\mathcal{S}_1) - \mathcal{S}_0\|^2) \sin\theta d\theta d\phi, \\ E_{\text{ext}} &= \int_0^{2\pi}\int_0^{\pi} (\lambda\|\mathcal{C}(\Phi(\mathcal{S}_1)) - \mathcal{C}(\mathcal{S}_0)\|^2) \sin\theta d\theta d\phi. \end{aligned}$$

Then this optimization procedure becomes a deformable model. We define the external force as

$$f_{\text{ext}} = -\nabla\mathcal{C}(\Phi(\mathcal{S}_1))(|\mathcal{C}(\Phi(\mathcal{S}_1)) - \mathcal{C}(\mathcal{S}_0)\|^2))$$

which penalizes the smoothness of the surfaces. Then by Davatzikos (1996), Φ is a smooth transformation which tends to preserve the relative positions of anatomical structures.

In this section, to illustrate our procedure, we are using more complex surfaces: the mandible surfaces. Two mandible surfaces are given: one is the template and the other is the surface-to-be-registered as shown in Figure 59. The matching transformations for the "Parallel Translation" of Gaussian and mean curvatures (Davatzikos, 1996) are shown in Figure 60. The iteration process of the registration is shown in Figure 61.

One may also formulate the registration problem in (46) using the elastic warping method, which is generalized from the elastic warping method from Ramsay and Li (1997). Let the warping function $h : [0, \pi] \times [0, 2\pi] \rightarrow [0, \pi] \times [0, 2\pi]$. This warping function has to be monotone so that the warping does not change the topology and the connection of the surfaces. Therefore, the warping

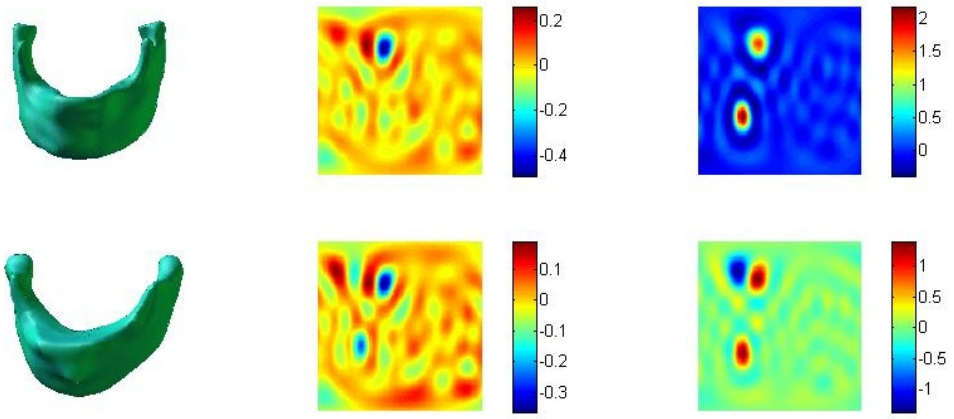

Figure 59: The surface-to-be-registered and its curvatures. The plots in first column are the two mandible surfaces; the plots in second column are the Gaussian curvatures; the plots in the third columns are the mean curvatures.

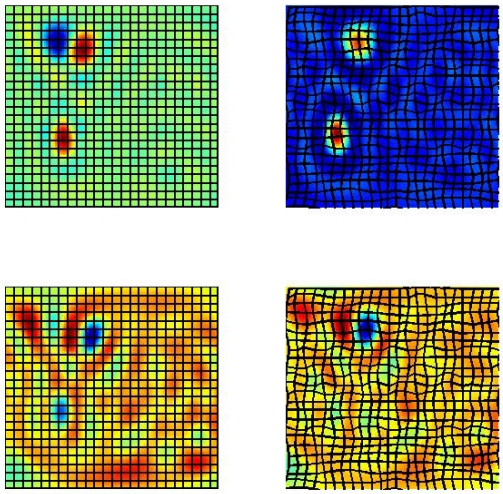

Figure 60: The plots in the first columns are the rectangle meshes on the Gaussian and mean curvature plots before registration; the plots in the second columns are the deformed rectangle meshes after non-linear registration.

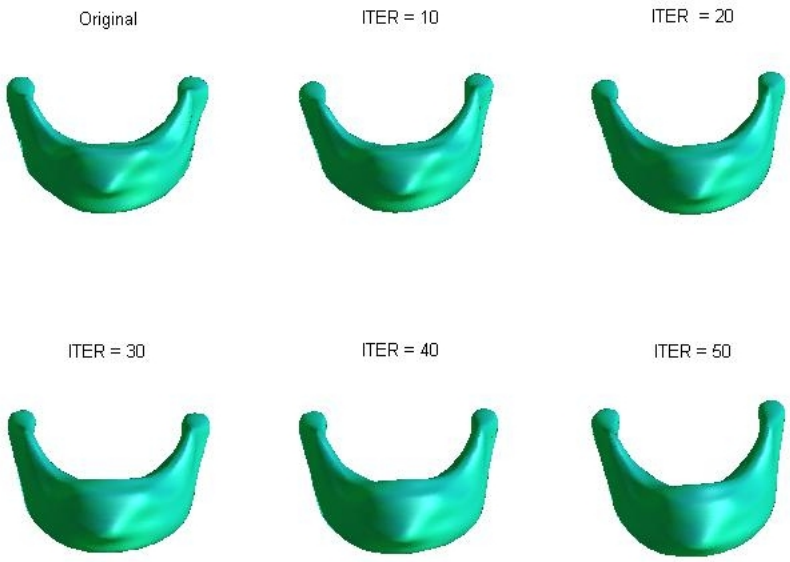

Figure 61: The iterative registration process of mandible surface in Figure 59.

function h minimizes

$$\int_0^{2\pi}\int_0^{\pi} \|\mathcal{S}_1(h(\theta,\phi)) - \mathcal{S}_2(\theta,\phi)\|^2 + \lambda \frac{\|\triangle h(\theta,\phi)\|}{\|\nabla h(\theta,\phi)\|} \sin\theta d\theta d\phi,$$

where ∇h is the gradient of h, and $\triangle h$ is the Laplacian of h. The tuning parameter can be estimated by generalized cross-validation (Wahba, 1990). For a given λ, the penalty on $1/(\nabla h)$ makes h monotone (its first derivatives are away from 0). The penalty on the Laplacian of h ensures the smoothness of h. So the penalty term yields both smoothness and monotonicity of the warping function.

The warping functions are usually constructed from a set of proper basis functions. Thin plate splines, were introduced to geometric design by Duchon

(1976). The theoretical details and numerical implementation can be found in Wahba (1990). The first and second derivatives of thin plates are smooth. The model of thin plate splines can be automatically tuned. It has closed-form solutions for both warping and parameter estimation (Wahba, 1990). Therefore, thin plate splines could be a good candidate for the warping functions.

Bibliography

Abell, F., Krams, M., Ashburner, J., Passingham, R., Friston, K., Frackowiak, R., Happe, F., Frith, C., Frith, U., 1999. The neuroanatomy of autism: a voxel-based whole brain analysis of structural scans. NeuroReport 10, 1647–1651.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Alexander, A., Lee, J., Lazar, M., Boudos, R., DuBray, M., Oakes, T., Miller, J., Lu, J., Jeong, E., McMahon, W., 2007. Diffusion tensor imaging of the corpus callosum in autism. Neuroimage 34 (1), 61–73.

Alley, W., 1987. A note on stagewise regression. The American Statistician 41 (2), 132–134.

Arfken, G., 1985. Development of the Fourier Integral, Fourier Transforms–Inversion Theorem, and Fourier Transform of Derivatives, 3rd edition. Academic Press, Florida.

Audette, M., Ferrie, F., Peters, T., 2002. An algorithmic overview of surface registration techniques for medical imaging. Medical Image Analysis 4 (3), 201–217.

Audette, M., Siddiqi, K., Ferrie, F., Peters, T., 2003. An integrated range-sensing, segmentation and registration framework for the characterization of intra-surgical brain deformations in image-guided surgery. Computer Vision and Image Understanding 89, 226–251.

Aupetit, B., 1991. A primer on spectral theory. Springer-verlag, New York.

Aylward, E., Minshew, N., Goldstein, G., Honeycutt, N., Augustine, A., Yates, K., Bartra, P., Pearlson, G., 1999. Mri volumes of amygdala and hippocampus in nonmentally retarded autistic adolescents and adults. Neurology 53, 2145–2150.

Baron-Cohen, S., Ring, H., Wheelwright, S., Bullmore, E., Brammer, M., Simmons, A., Williams, S., 1999. Social intelligence in the normal and autistic brain: an fmri study. Eur J Neurosci. 11, 1891–1898.

Becker, R., Chambers, J., Wilks, A., 1988. The S Language. Wadsworth and Brooks/Cole.

Beg, M., Miller, M., Trouve, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision 61 (2), 139–157.

Berezankii, J., 1968. Expansions in Eigenfunctions of Self-adjoint Operators. American Mathematical Society, ISBN 0821815679.

Berument, S., Rutter, M., Lord, C., Pickles, A., Bailey, A., 1999. Autism screening questionnaire: diagnostic validity. British Journal of Psychiatry 175, 444–451.

Besl, P., Jain, R., 1986. Segmentation through variable-order surface fitting. Computer Vision, Graphics and Image Process 33, 86–91.

Bickel, P., 2007. Discussion: The dantzig selector: Statistical estimation when p is much larger than n. Annals of Statistics 35 (6), 2352–2357.

Bickel, P., Doksum, K., 2000. Mathematical Statistics: Basic Ideas and Selected Topics. Prentice Hall, Upper Saddle River, NJ.

Bluestein, L., 1968. A linear filtering approach to the computation of the discrete fourier transform. Northeast Electronics Research and Engineering Meeting Record 10, 218–219.

Bonvallet, B., Griffin, N., Li, J., 2007. A 3d shape descriptor: 4d hyperspherical harmonics. Proceedings of the 2007 IASTED International Conference on Graphics and Visualization in Engineering, 113–116.

Bookstein, F., 1997. Shape and the information in medical images: a decade of the morphometric synthesis. Comp. Vision and Image under. 66 (2), 97–118.

Bosi, M., Goldberg, R., 2003. Introduction to Digital Audio Coding and Standards. Kluwer Academic Publishers, Boston.

Bracewell, R., 1999. The Fourier Transform and Its Applications, third edition. McGraw-Hill Book Co., New York.

Brandenburg, K., Bosi, M., 1997. Overview of mpeg audio: Current and future standards for low-bit-rate audio coding. Journal of the Audio Engineering Society 45, 4–21.

Brechbuehler, C., Gerig, G., Kuebler, O., 1995. Parametrization of closed surfaces for 3d shape description. Comp. Vision and Image Underst. (CVIU) 61 (2), 154–170.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees. Wadsworth.

Bro-Nielsen, M., Gramkow, C., 1996. Fast fluid registration of medical images. Lecture Notes in Computer Science 1131, 267–276.

Bronstein, M., Bronstein, A., Zibulevsky, M., Azhari, H., 2002. Reconstruction in diffraction ultrasound tomography using nonuniform fft. Medical Imaging, IEEE Transactions on 21 (11), 1395–1401.

Bruun, G., 1978. z-transform dft filters and ffts. IEEE Trans. on Acoustics, Speech and Signal Processing 26 (1), 56–63.

Bulow, T., 2004. Spherical diffustion for 3d surface smoothing. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 1650–1654.

Burnham, K., Anderson, D., 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag. New York.

Byerly, W., 1959. An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics. New York: Dover.

Candes, E., Tao, T., 2005. The dantzig selector: statistical estimation when p is much larger than n.

Cao, J., Worsley, K., 1999. The detection of local shape changes via the geometry of hotelling's t^2 fields. Annals of Statistics 27, 925–942.

Casey, J., 1996. Exploring Curvature. Vieweg: Germany.

Cauchy, A., 1842. Comptes Rend 15.

Chambers, J., Hastie, T., 1992. Statistical Models in S. Wadsworth and Brooks/Cole.

Choudhury, R., Fuster, V., Badimon, J., Fisher, E., Fayad, Z., 2002. Mri and characterization of atherosclerotic plaque: Emerging applications and molecular imaging. Arterioscler. Thromb. Vasc. Biol. 22, 1065–1074.

Chowning, J., 1973. The synthesis of complex audio spectra by means of frequency modulation. Journal of the Audio Engineering Society 21 (7), 526–534.

Chung, M., 2006a. Heat kernel smoothing on unit sphere. IEEE International Symposium on Biomedical Imaging (ISBI) 1430.

Chung, M., 2006b. Heat kernel smoothing on unit sphere. IEEE International Symposium on Biomedical Imaging 1430.

Chung, M., Dalton, K., Alexander, A., Davidson, R., 2004. Less white matter concentration in autism: 2d voxel-based morphometry. NeuroImage 23, 242–251.

Chung, M., Dalton, K., Davidson, R., 2008a. Tensor-based cortical surface morphometry via weighted spherical harmonic representation. IEEE transactions on medical imaging (in press).

Chung, M., Dalton, K., Shen, L., Evans, A., Davidson, D., 2007a. Wieghted fourier series representation and its application to quantifying the amount of gray matter. IEEE Transaction on Medical Imaging 26 (4), 566–581.

Chung, M., Hartley, R., Dalton, K., Davidson, R., 2007b. Encoding cortical surface by spherical harmonics. Statistics Sonica (in press).

Chung, M., Nacewicz, B., Wang, S., Dalton, K., Pollak, S., Davidson, R., 2008b. Amygdala surface modeling with weighted spherical harmonics. submitted. MIAR 2008 (in press).

Chung, M., Robbins, S., Dalton, K., Davidson, R., Alexander, A., Evans, A.,

2005. Cortical thickness analysis in autism with heat kernel smoothing. NeuroImage 25, 1256–1265.

Chung, M., Robbins, S., Dalton, K., Wang, S., Evans, A., Davidson, R., 2006a. Tensor-based cortical morphometry via weighted spherical harmonic representation. IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA).

Chung, M., Shen, L., Dalton, K., Davidson, D., 2006b. Multi-scale voxel-based morphometry via weighed spherical harmonic representation. Lecture Notes in Computer Science (LNCS) 4091, 36–43.

Collins, G., 2004. The shapes of space. Sci. Amer. 291, 94–103.

Convit, A., McHugh, P., Wolf, O., de leon, M., Bobinikski, M., De Santi, S., Roche, A., Tsui, W., 1999. Mri volume of the amygdala: a reliable method allowing separation from the hippocampal infomation. Psychiatry Res. 90, 113–123.

Conway, J., 1985. A course in functional analysis. Springer Verlag.

Cooley, J., Tukey, J., 1965. An algorithm for the machine calculation of complex fourier series. Math. Comput. 19, 297–301.

Courant, R., Hilbert, D., 1953. Methods of mathemaical physics. Wiley, New York.

Coxter, H., 1969. Introduction to Geometry, 2nd editiion. New York: Wiley.

Dalton, K., Nacewicz, B., Johnstone, T., Schaefer, H., Gernsbacher, M., Goldsmith, H., Alexander, A., Davidson, R., 2005a. Gaze fixation and the neural circuitry of face processing in autism. Nat. Neurosci. 8 (4), 519–526.

Dalton, K., Nacewicz, B., Johnstone, T., Schaefer, H., Gernsbacher, M., Goldsmith, H., Alexander, A., Davidson, R., 2005b. Gaze fixation and the neural circuitry of face processing in autism. Nat Neurosci. 8, 519–526.

Davatzikos, C., 1996. Nonlinear registration of brain images using deformable models. Proc. of the IEEE Workshop on Math. Methods in Biomedical Image Analysis.

de Boor, C., 1978. A Practical Guide to Splines. New York: Springer-Verlag.

De Nicolao, G., Sparacino, G., CoBelli, C., 1997. Nonparametric input estimation in the physiological system: problems, methods and case studies. Automatica 5, 851–870.

doCarmo, M., 1976. Differential Geometry of Curves and Surfaces. Prentice Hall.

Dragomir, S., 2006. Differential geometry and analysis on CR manifold. Boston: Birkhauser.

Duchon, J., 1976. Splines minimizing rotation invariant seminorms in sobolev spaces. Constructive Theory of Functions of Several Variables 1, 85–100.

Duren, P., Hengartner, W., 1997. Harmonic mappings of multiply connected domains. Pac. J. Math. 180, 201–220.

Eaton, R. P., Allen, R. C., Schade, D. S., Erickson, K. M., Standefer, J., 1980. Prehepatic insulin production in man: Kinetic analysis using peripheral connecting peptide behavior. J. Clin. Endocrinol. Metab. 51, 520–528.

Eck, M., DeRose, T., Duchamp, T., Hoppe, H., Lounsbery, M., Stuetzle, W., 1995. Multiresolution analysis of arbitrary meshes. Proceedings of SIGGRAPH, 173–182.

Efron, B., Hastie, T., Tibshirani, R., 2007. Discussion of the dantzig selector.

Elliott, D., Rao, K., 1982. Fast Transforms: Algorithms, Analyses, and Applications. Academic Press: New York.

Fan, T.J. adn Medioni, G., Nevatia, R., 1986. Description of surfaces from range data using curvature properties. Proc. Comput. Vision Patt. Recogn., 86–91.

Fan, J., Wang, M., Yao, Q., 2008. Modelling multivariate volatilities via conditionally uncorrelated components. Journal of Royal Statistical Society B, to appear.

Fischer, B., Modersitzki, J., 2004. A unified approach to fast image registration and a new curvature based registration technique. Linear Algebra and its Applications 380, 107–124.

Forster, M., 2000. Key concepts in model selection: Performance and generalizability. Linear Algebra and its Applications 44, 205–231.

Fourier, J., 1822. Théorie analytique de la chaleur.

Frank, R., Hargreaves, R., 2003. Clinical biomarkers in drug discovery and development. Nature Reviews Drug Discovery 2, 566–580.

Freund, R., Vail, R., Clunies-Ross, C., 1961. Residual analysis. Journal of American Statistical Association 56, 98–104.

Frigo, M., Johnson, S., 2005. The disign and implementation of fftw3. Proceeding of the IEEE 93 (2), 216–231.

Gefen, S., Tretiak, O., Nissanov, J., 2003. Elastic 3-d alignment of rat brain histological images. IEEE TRANSACTIONS ON MEDICAL IMAGING 22 (11), 1480–1489.

Gerig, G., Styner, M., Jones, D., Weinberger, D., Lieberman, 2001. Shape analysis of brain ventricles using spharm. MMBIA, 171–178.

Gerig, G., Styner, M., Szekely, 2002. Statistical shape models for segmentation and structural analysis. Proc. IEEE Int. Symp. Biomed. Imag. (ISBI), 18–21.

Goldberger, A., 1961. Stepwise least squares: residual analysis and specification error. Journal of American Statistical Association 56, 998–1000.

Goldberger, A., Jochemes, D., 1961. Note on stepwise least squares. Journal of American Statistical Association 56, 105–110.

Golland, P., Grimson, W., Kikinis, R., 1999. Statistical shape analysis using fixed topology skeletons: Corpus callosum study. IPMI LNCS 1613, 382–388.

Gonzalez, O., Maddocks, J., 1996. Global curvature, thickness and ideal shapes of knots. The Proceedings of the National Academy of Sciences, USA 96, 4767–4773.

Good, I., 1958. The interaction algorithm and practical fourier analysis. Journal of the Royal Statistical Society, Series B 20 (2), 361–371.

Gorbachuk, M., 1998. Operator approach to the cauchy-kovalevskaya thoerem. Journal of Mathematical Sciences 99 (5), 1527–1532.

Gotsman, C., Gu, X., Sheffer, A., 2003. Fundamentals of spherical parameterization for 3d meshes. ACM Transactions on Graphics 22, 358–363.

Gottlieb, D., Gustafsson, B., Forssen, P., 2000. On the direct fourier method for computer tomography. Medical Imaging, IEEE Trans. on 19 (3), 223–232.

Gottlieb, D., Shu, C., 1997. On the gibbs phenomenon and its resolution. SIAM Review 39 (4), 644–668.

Gray, A., 1997. Modern Differential Geometry of Curves and Surfaces with Mathematica, 2nd ed. Boca Raton, FL: CRC Press.

Greengard, L., 1994. Fast algorithms for classical physics. Science 265, 909–914.

Groemer, H., 1996. Geometric Applications of Fourier Series and Shperical Harmonics. Cambridge University Press, New York.

Gu, X., Wang, Y., Chan, T., Tompson, T., Yau, S., 2004. Genus zeros surface conformal mapping and its application to brain surface mapping. IEEE Trans. Med. Imag. 20 (8), 1–10.

Halmos, P., 1978. Measure theory. Springer Verlag.

Hardan, A., Minshew, N., Keshavan, M., 2000. Corpus callosum size in autism. Neurology 55, 1033–1036.

Harris, F., 1978. On the use of windows for harmonic analysis with the discrete fourier transform. Proceedings of the IEEE 66, 51–83.

Hawkins, W., 1996. Fourier transform resampling: theory and application. Nuclear Science Symposium, 1996. Conference Record., 1996 IEEE 3, 1491–1495.

Haznedar, M., Buchsbaum, M., Wei, T., Hof, P., Cartwright, C., Bienstock, C., Hollander, E., 2000. Limbic circuitry in patients with autism spectrum disorders studies with positron emission tomography and magnetic resonance imaging. American Journal of Psychiatry 157, 1994–2001.

Healy, D., Rockmore, D., Kostelec, P., Moore, S., 2003. Ffts for the 2-sphere - improvements and variations. The Journal of Fourier Analysis and Applications 9 (4), 341–385.

Hobson, E., 1955. The Theory of Spherical and Ellipsoidal Harmonics. Chelsea, New York.

Hocking, R., 1976. The analysis and selection of variables in linear regression. Biometrics 32, 321–331.

Hoffmann, T., Chung, M., Dalton, K., Alexander, A., Wahba, G., Davidson, R., 2004. Subpixel curvature estimation of the corpus callosum via splines and its application to autism. 10th Annual Meeting of the Organization for Human Brain Mapping.
URL http://www.stat.wisc.edu/ mchung/papers/HBM2004/HBM2004thomas.html.

Horn, R., Johnson, C., 1985. Matrix Analysis. Cambridge University Press, London.

Hovorka, R., Chappell, M., Godfrey, K., Madde, F., Rouse, M., Soons, P., 1998. Code: A deconvolution program implementing a regularization method of deconvolution consgtrained to non-nagetive values. design and pilot evaluation. Biopharm. Drug Dispos. 19, 39–53.

Hsiung, C., 1981. A First Course in Differential Geometry. John Wiley and Sons, New York.

Jost, J., 2002. Riemannian Geometry and Geometric Analysis. Springer-Verlag, Berlin.

Kass, M., Witkin, A., Terzopoulos, D., 1987. Snakes: active contour models. International Journal of Computer Vision 1 (4), 321–331.

Kazhdan, M., Funkhouser, T., Rusinkiewicz, S., 2003. Rotation invariant spherical harmonic representation of 3d shape descriptors. In: Symposium on Geometry Processing.

Kelemen, A., Szekely, G., Gerig, G., 1999. Elastic model-based segmentation of 3d neuroradiological data sets. IEEE Transactions on Medical Imaging 18, 828–839.

Kiebel, S. J., Poline, J., Friston, K., Holmes, A., Worsley, K., 1999. Robust smoothness estimation in statistical parametric maps using standarized residuals from the general linear model. NeuroImage 10, 756–766.

Kim, H., Loh, W.-Y., 2001. Classification trees with unbiased multiway splits. Journal of the American Statistical Association 96, 589–604.

Klette, R., Rosenfeld, A., 2004. Digital Geometry. Morgan Kaufmann: San Francisco.

Kowalevski, S., 1875. Zur theorie der partiellen differentialgleichung. Journal fur die reine und angewandte Mathematik 80, 1–32.

Krantz, S., 1999. Handbook of complex variables. Birkhuser.

Kreyszig, E., 1991. Principal Normal, Curvature, Osculating Circle. Dover, New York.

Kuhnel, W., 2000. Differential Geometry: Curves-Surfaces-Manifolds. American Mathematics Association.

Lawrence, J., 1972. A Book of Curves. New York: Dover.

Lee, Y., Kim, H., Lee, S., 2002. Mesh parameterization with a virtual boundary. Computers and Graphics (Special Issue of the 3rd Israel-Korea Binational Conf. on Geometric Modeling and Computer Graphics) 26 (5), 677–686.

Levy, B., Mallet, J., 1998. Non-distorted texture mapping for sheared triangulated meshes. Proceedings of SIGGRAPH, 343–352.

Leymarie, F., Levine, M., 1993. Tracking deformable objects in the plane using an active contour model. IEEE Trans. on Pattern Anal. Machine Intell. 15 (6), 617–634.

Lipschutz, S., Lipson, M., 2001. Schaum's Outlines: Linear Algebra. Tata McGraw-hill edition: Delhi.

Lockwood, E., 1961. A Book of Curves. Great Britian: Cambridge University Press.

Loh, W., 2002. Regression trees with unbiased variable selection and interaction detection. Statistics Sinica 12, 361–368.

Loh, W., Shih, Y., 1997. Split selection methods for classification trees. Statistics Sinica 7, 815–840.

Loh, W., Vanichsetaku, N., 1988. Tree-structured classification via generalized discriminant analysis (with discussion). Journal of the American Statistical Association 83, 715–728.

Lorensen, W., Cline, H., 1987. Marching cubes: A high resolution 3d surface construction algorithm. Computer Graphics 21 (4).

Lustig, M., Tsaig, J., Lee, J. H., Donoho, D., 2004. Fast spiral fourier transform for iterative mr image reconstruction. IEEE International Symposium on Volume 1, 15–18.

Mallat, S., Zhang, Z., 1993. Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing 41, 3397–3415.

Mallow, C., 1973. Some comments on cp. Technometrics 15, 661–675.

Martyna, G., Berne, B., 1989. Structure and energies of $x_{e_n^-}$, many body polarization effects. J. Chem. Phys. 90 (7), 3744–3755.

Matej, S., Bajla, I., 1990. A high-speed reconstruction from projections using direct fouriermethod with optimized parameters-an experimental analysis. Medical Imaging, IEEE Transactions on 9 (4), 421–429.

Matheny, A., Goldgof, D., 1995. The use of three- and four-dimensional surface harmonics for rigid and nonrigid shapce recoverary and represenation. IEEE Trans. on Pattern Analysis and Machine Intelligence 17 (10), 967–981.

McInerney, T., Terzopoulos, D., 1996. Deformable models in medical image analysis: a survey. Medical Image Analysis 1 (2), 91–108.

McKeague, I., 2005. A statistical model for signiture verification. Journal of the American Statistical Association 100, 231–241.

Mezrich, R., 1995. A perspective on k-space. Radiology 195, 297–315.

Miller, M., Joshi, S., Maffitt, D., McNally, J., Grenander, U., 1994. Membranes, mitochondria and amoebe: shape models. Advances in applied statistics, 137–159.

Mitchell, R., Littlejohn, R., 1997. Derivation of planar three-body hyperspherical harmonics. Physics Review 56.

Morosov, V., 1966. On the solution of functional equations by the method of regularization. Soviet Math. Dokl. 7, 414–423.

Morosov, V., 1984. Methods for solving incorrectly posed problems. Springer-Verlag.

Nacewicz, B., Dalton, K., Johnstone, T., Long, M., McAuliff, E., Oakes, T., Alexander, A., Davidson, R., 2006. Amygdala volume and nonverbal social impairment in adolescent and adult males with autism. Archives of General Psychiatry 63, 1417–1428.

Nakhushev, A., 2001. Cauchy-kovalevskaya theorem. Encyclopaedia of Mathematics 978.

Okunev, P., Johnson, C., 1997. Necessary And Sufficient Conditions For Existence of the LU Factorization of an Arbitrary Matrix. Numerical Analysis, arXiv:math/0506382v1.

Osborne, M., Presnell, B., Turlach, B., 2000. A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis 20, 389–404.

Page, D., Sun, Y., Koschan, F., Paik, J., Abidi, M., 2002. Normal vector voting: Crease detection and curvature estimation on large, noisy meshes. Graphical Models 64, 199–229.

Pien, H., Fischman, A., Thrall, J., Sorensen, A., 2005. Using imaging biomarkers to accelerate drug development and clinical trials. Drug Discovery Today 10 (4), 259–266.

Pierce, K., Muller, R., Ambrose, J., Allen, G., Courchesne, E., 2001. Face processing occurs outside the fusiform face area in autism: evidence from functional mri. Brain 124, 2059–2073.

Piven, J., Bailey, J., Ranson, B., Arndt, S., 1997. An mri study of the corpus callosum in autism. Am. J. Psychaitry 154 (8), 1051–1056.

Rader, C., 1968. Discrete fourier transforms when the number of data samples is prime. Proc IEEE 56, 1107–1108.

Ramsay, J., Li, X., 1997. Curve registration. J. R. Statist. Soc. B 60 (2), 351–363.

Ramsay, J., Silverman, B., 1997. Functional Data Analysis. New York: Springer-Verlag.

Ramsay, J., Silverman, B., 2002. Applied Functional Data Analysis. New York: Springer-Verlag.

Robinson, S., 2003. Russian reports he has solved a celebrated math problem. New York Times 3.

Rosenberg, S., 1997. The Laplacian on a Riemannian Manifold. Cambridge University Press.

Rowe, D., 2005. Modeling both magnitude and phase of complex-valued fmri data. NeuroImage 25, 1310–1324.

Rowe, D., Logan, B., 2004. A complex way to computefmri activation. NeuroImage 24, 1078–1092.

Rowe, D., Nencka, A., Hoffman, R., 2007. Signal and noise of fourier reconstructed fmri data. Journal of Neuroscience Methods 159, 361–369.

Rubin, W., 1991. Functional Analysis. McGraw-Hill.

Rudin, W., 1976. Principles of mathematical analysis. McGraw-Hill, New York.

Sander, P., Zucker, S., 1986. Stable surface estimation. Proc. Intl Conf. Patt. Recogn. 1, 1165–1167.

Scammon, R., 1930. The measurement of the body in childhood. Minneapolis: University of Minnesota Press.

Schomberg, H., Timmer, J., 1995. The gridding method for image reconstruction by fouriertransformation. Medical Imaging, IEEE Transactions on 14 (3), 596–607.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6 (2), 461–464.

Scott, F., Baron-Cohen, S., Bolton, P., Brayne, C., 2002. The cast (childhood asperger syndrome test): preliminary development of a uk screen for mainstream primary-school-age children. Autism 2 (1), 9–31.

Shao, J., 2003. Mathematical Statistics. Springer-New York.

Shen, L., Chung, M., 2006. Large-scale modeling of parametric surfaces using spherical harmonics. Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT).

Shen, L., Ford, J., Makedon, F., Saykin, A., 2004. Surface-based approach for classificaion of 3-d neuroanatomical structures. Intell. Data Anal. 9, 519–542.

Shi, P., Robinson, G., Duncan, J., 1994. Myocardial motion and function assessment using 4d images. Proc. IEEE Conf. Vis. Biomedical Comput.

Silverman, B., 1995. Incorporating parametric effects intro functional principle component analysis. Journal of the Royal Statistical Society, Series B 57, 673–698.

Sparacino, G., Pillonetto, G., Capello, M., De Nicalao, G., Cobelli, C., 2001. Winstodec: a stochastic deconvolution interactive program for physiological and pharmacokinetic systems. Computer methods and programs in biomedicine 67, 67–77.

Sparks, B., Friedman, S., Shaw, D., Aylward, E., Echelard, D., Artru, A., Maravilla, K., Giedd, J., Munson, J., Dager, S., 2002. Brain structural abnormalities in young children with autism spectrum disorder. Neurology 59, 184–192.

Sternberg, W., Smith, T., 1946. The Theory of Potential and Spherical Harmonics, 2nd ed. Toronto: University of Toronto Press.

Stevens, K., 1981. Computer Vision. Noth Holland Publishing Company: Amsterdam.

Stoker, J., 1969. Differential geometry. Wiley-New York.

Strang, G., 2003. Introduction to Linear Algebra, 3rd edition. Wellesley, Massachusetts: Wellesley-Cambridge Press.

Styner, M., Oguz, I., Xu, S., Brechbuhler, C., Pantazis, D., Levitt, J., Shenton,

M., Gerig, G., 2006. Framework for the statistical shape analysis of brain structures using spharm-pdm. Insight J., 1–20.

Taguchi, K., Zeng, G., Gullberg, G., 2001. Cone-beam image reconstruction using spherical harmonics. Phys. Med. Biol. 46, 127–138.

Tang, X., 2005. A sampling framework for accurate curvature estimation in discrete surfaces. IEEE Transactions on Visualization and Computer Graphics 11 (5), 573–583.

Taylor, J., Worsley, K., 2007. Random fields of multivariate test statistics, with applications to shape analysis. Annals of Statistics, accepted.

Terzopoulos, D., Fleischer, K., 1988. Deformable models. The Visual Computer 4, 306–331.

Tibshirani, R., 1996. Regression shrinkage and selection via lasso. Journal of Royal Statistical Society, Series B (Methodological) 58 (1), 267–288.

Toffolo, G., Breda, E., Cavaghan, M., Ehrman, D., Polonsky, K., Cobelli, C., 2001. Quantitative indexes of cell function during graded up and down glucose infusion from c-peptide minimal models. Am. J. Physiol. Endocrinol. Metab. 280, E2–E20.

Tong, W., Tang, C., 2005. Robust estimation of adaptive tensors of curvature by tensor voting. IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3), 434–449.

Toponogov, V., 2006. Differential Geometry of Curves and Surfaces. Birkhauser: Boston.

Trott, M., 2004. The Mathematica GuideBook for Programming. Springer-Verlag, New York.

Vemuri, B., Mitiche, A., Aggarwal, J., 1986. Curvaure-based representation of objects from range data. Image and Vision Computing 4 (2), 107–114.

Vidal, C., DeVito, T., Hayashi, K., Drost, D., Williamson, P., Craven-Thuss, B., Herman, D., Sui, Y., Toga, A., Nicolson, R., Thompson, P., 2003. Detection and visualization of corpus callosum deficits in autistic children using novel anatomical mapping algorithms,. Proc. International Society for Magnetic Resonance in Medicine.
URL http://www.loni.ucla.edu/ thompson/ISMRM2003/cvISMRM2003.html

Viola, P., Wells, W., 1995. Alignment by maximization of mutual information. Fifth International Conference on Computer Vision, IEEE, 16–23.

vonSeggern, D., 1994. Practical Handbook of Curve Design and Generation. CRC Press, Inc.

Vorperian, H., Durtschi, R., Wang, S., Chung, M., Ziegert, A., Gentry, L., 2006. Estimated head circumference from imaging studies. Journal of Radiology, accepted.

Vorperian, H., Kent, R., Gentry, L., Yandell, B., 1999. Mri procedures to study the concurrent anatomic development of the vocal tract structures: Preliminary results. International Journal of Pediatric Otorhinolaryngology 49 (3), 721–736.

Vorperian, H., Kent, R., Lindstrom, M., Kalina, C., Gentry, L., Yandell, B., 2005. Development of vocal tract length during early childhood: A magentic resonance imaging study. Journal of the Acoustical Society of America 117 (1), 721–736.

Wahba, G., 1990. Spline models for observational data. SIAM.

Waiter, G., Williams, J., Murray, A., Gilchrist, A., Perrett, D., Whiten, A., 2005. Structural white matter deficits in high-functioning individuals with autistic spectrum disorder: a voxel-based investigation. NeuroImage 24 (2), 455–461.

Wang, S., 2003. Numerical approximation of $c^{1,1}$-curves. Master Thesis.

Weisstein, E., 2002. Poincare conjecture purported proof perforated. MathWorld Headline News.

Worsley, K., 1996. An unbiased estimator for the roughness of a multivariate gaussian random field. Technical report.

Worsley, K., 2001. Testing for signals with unknown location and scale in a

chi-squared random field, with an application to fmri. Advances in Applied Probability 33, 773–793.

Worsley, K., Marrett, S., Neelin, P., Evans, A., 1995. A unified statistical approach for determining significant signals in location and scale space images of cerebral activation. Quantification of brain function using PET.

Wu, H., Barba, J., Gil, J., 1996. An iterative algorithm for cell segmentation using short-time fourier transform. J. Microsc 184 (2), 127–132.

Xu, C., 1999. Deformable models with application to human cerebral cortex reconstruction from magnetic resonance images. Ph.D Thesis, John Hopkins University.

Xu, C., Prince, J., 1997. Snakes, shapes, and gradient vector flow. IEEE Transactions on Image Processing 7 (3), 359–369.

Yeargin-Allsopp, M., C., R., Karapurkar, T., Doernberg, N., Boyle, C., Murphy, C., 2003. Prevalence of autism in a us metropolitan area. The Journal of American Medical Association 289 (1), 49–55.

Yeo, B., 2005. Computing spherical transform and convolution on the 2-sphere. Manuscript, MIT.

Zigelman, G., Kimmel, R., Kiryati, N., 2001. Texture mapping using surface flattening via multi-dimensional scaling. IEEE Trans. Visualization and Computer Graphics 8 (2), 198–207.

Zwicker, E., Fastl, H., 1999. Psychoacoustics: Facts and Models. Springer Verlag, Berlin.