Moo K. Chung

Statistical Methods in Brain Image Analysis with MATLAB

1

General Linear Models

In this chapter, we cover *general linear models* (GLM) that have been widely used in brain imaging applications. The GLM is a very flexible and general statistical framework encompassing a wide variety of fixed effect models such as the multiple regressions, the analysis of variance (ANOVA), the multivariate analysis of variance (MANOVA), the analysis of covariance (ANCOVA) and the multivariate analysis of covariance (MANCOVA) [119]. Note that the term *linear* is misleading in a sense that the model can also include mathematically nonlinear model terms such as the higher degree polynomials.

The GLM provides a framework for testing various associations and hypotheses while accounting for nuisance covariates in the model in a straightforward fashion. The effect of age, sex, brain size and possibly IQ can have severe confounding effects on the final outcome of many anatomical and functional imaging studies. Older population's reduced functional activation could be the consequence of age-related atrophy of neural systems [84]. Brain volumes is significantly larger for children with autism 12 years old and younger compared with normally developing children [11]. Therefore, it is desirable to account for various confounding factors such as age and sex in the model. This can be done using GLM. The parameters of the GLM are mainly estimated by the least squares estimation and has been implemented in many statistical packages such as R (`www.r-project.org`) or Splus [92] and brain imaging packages such as SPM (`www.fil.ion.ucl.ac.uk/spm`) and fMRI-STAT (`www.math.mcgill.ca/keith/fmristat`).

1.1 General Linear Models

Let y_i be the response variable, which is mainly coming from images and $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})$ to be the variables of interest and $\mathbf{z}_i = (z_{i1}, \cdots, z_{ik})$ to be nuisance variables corresponding to the i-th subject. We assume there are n subjects. We are interested in testing the significance of the group variable while accounting for age and gender. In a more general setting, we have a GLM

$$y_i = \mathbf{z}_i \boldsymbol{\lambda} + \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i$$

where $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_k)'$ and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ are unknown parameter vectors to be estimated. We assume ϵ to be the usual zero mean Gaussian noise.

The significance of the variable of interests $\mathbf{x}_i$ is determined by testing the null hypothesis

$$H_0 : \boldsymbol{\beta} = 0 \text{ vs. } H_1 : \boldsymbol{\beta} \neq 0.$$

The fit of the reduced model corresponding to $\beta = 0$, i.e.

$$y_i = \mathbf{z}_i \boldsymbol{\lambda}, \tag{1.1}$$

is measured by the sum of the squared errors (SSE):

$$\text{SSE}_0 = \sum_{i=1}^{n} (y_i - \mathbf{z}_i \widehat{\boldsymbol{\lambda}}_0)^2,$$

where $\widehat{\boldsymbol{\lambda}_0}$ is the least squares estimation obtained from the reduced model. The reduced model (1.1) can be written in a matrix form

$$\underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} z_{11} & \cdots & z_{1k} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nk} \end{pmatrix}}_{\mathbf{Z}} \underbrace{\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}}_{\boldsymbol{\lambda}}. \tag{1.2}$$

By multiplying $\mathbf{Z}'$ on the both sides, we obtain

$$\mathbf{Z}'\mathbf{y} = \mathbf{Z}'\mathbf{Z}\boldsymbol{\lambda}.$$

Now the matrix $\mathbf{Z}'\mathbf{Z}$ is a full rank and can be invertible if $n \geq k$, which is the usual case in brain imaging. Therefore, the matrix equation can be solved by performing a matrix inversion

$$\widehat{\boldsymbol{\lambda}}_0 = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}.$$

Similarly the fit of the full model corresponding to $\boldsymbol{\beta} \neq 0$, i.e.

$$y_i = \mathbf{z}_i \boldsymbol{\lambda} + \mathbf{x}_i \boldsymbol{\beta} \tag{1.3}$$

is measured by

$$\texttt{SSE}_1 \quad = \quad \sum_{i=1}^{n} (y_i - \mathbf{z}_i \widehat{\boldsymbol{\lambda}}_1 - \mathbf{x}_i \widehat{\boldsymbol{\beta}}_1)^2,$$

where $\widehat{\boldsymbol{\lambda}}_1$ and $\widehat{\boldsymbol{\beta}}_1$ are the least squares estimation from the full model. The full model can be written in a matrix form by concatenating the row vectors $\mathbf{z}_i$ and $\mathbf{x}_i$ into a larger row vector $(\mathbf{z}_i, \mathbf{x}_i)$, and the column vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ into a larger column vector $(\boldsymbol{\lambda}', \boldsymbol{\beta}')'$. Then the full model can be also written in a matrix form and solved by the matrix inversion.

Note that

$$\texttt{SSE}_1 = \min_{\boldsymbol{\lambda}_1, \boldsymbol{\beta}_1} \sum_{i=1}^{n} (y_i - \mathbf{z}_i \boldsymbol{\lambda}_1 - \mathbf{x}_i \boldsymbol{\beta}_1)^2 \leq \min_{\boldsymbol{\lambda}_0} \sum_{i=1}^{n} (y_i - \mathbf{z}_i \boldsymbol{\lambda}_0)^2 = \texttt{SSE}_0.$$

So the larger the value of $\texttt{SSE}_0 - \texttt{SSE}_1$, more significant the contribution of the coefficients $\boldsymbol{\beta}$ is. Under the assumption of the null hypothesis H_0, the test statistic is the ratio

$$F = \frac{(\texttt{SSE}_0 - \texttt{SSE}_1)/p}{\texttt{SSE}_0/(n-p-k)} \sim F_{p,n-p-k}. \tag{1.4}$$

The larger the F value, it is more unlikely to accept H_0.

When $p = 1$, the test statistic F is distributed as $F_{1,n-1-k}$, which is the square of the student t-distribution with $n - 1 - k$ degrees of freedom, i.e. t^2_{n-1-k}. In this particular case, it is better to use t-statistic. The advantage of using the t-statistic is that unlike the F-statistic, it has two sides so we can actually use it to test for one sided alternative hypothesis $H_1 : \beta_1 \geq 0$ or $H_1 : \beta_1 \leq 0$. Therefore, the t-statistic map can provides the direction of the group difference that the F-statistic map cannot provide.

1.1.1 R-square

The R-square of a model explains the proportion of variability in measurement that is accounted by the model. Sometime R-square is called the coefficient of determination and it is given as the square of a correlation coefficient for a very simple model. For a linear model involving the response variable y_i, the total sum of squares (SST) measures total total variation in response y_i and is defined as

$$\texttt{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2,$$

where is the sample mean of y_i.

On the other hand, SSE measures the amount of variability in y_i that is not explained by the model. Note that SSE is the minimum of the sum of squared residual of any linear model, SSE is always smaller than SST. Therefore, the amount of variability explained by the model is SST-SSE. The proportion of variability explained by the model is then

$$R^2 = \frac{\texttt{SST} - \texttt{SSE}}{\texttt{SST}},$$

which is the coefficient of determination. The R-square ranges between 0 and 1 and the value larger than 0.5 is usually considered as significant.

1.1.2 GLM for Whole Brain Images

In brain imaging, the linear model of type (3.16) is usually fitted in each voxel separately. If image dimension is of size $100 \times 100 \times 200$ for instance, we need to fit 2 million linear models, which causes a serious computational bottleneck. So what we need is to reformulate the problem such that we fit all linear models simultaneously in a slice so that we only need to perform the least squares estimation 200 times.

Let $\mathbf{y}_j$ be the measurement vector at the j-th voxel in a slice. Assume there are m voxels in a slice. We have the same design matrix $\mathbf{Z}$ for all m voxels. Then we need to estimate the parameter vector $\boldsymbol{\lambda}_j$ in

$$\mathbf{y}_j = \mathbf{Z}\boldsymbol{\lambda}_j. \tag{1.5}$$

each j. Instead of solving (1.6) separately, we combine all of them together so that we have matrix equation

$$\underbrace{[\mathbf{y}_1, \cdots, \mathbf{y}_m]}_{\mathbf{Y}} = \mathbf{Z}\underbrace{[\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_m]}_{\boldsymbol{\Lambda}}. \tag{1.6}$$

The least squares estimation of the parameter matrix $\boldsymbol{\Lambda}$ proceeds similarly and given by

$$\widehat{\boldsymbol{\Lambda}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}.$$

The least squares estimation technique does not work for sparsely sampled data where $n \ll k$. In this case, $\mathbf{Z}'\mathbf{Z}$ is size $k \times k$ but only of rank n. So we can't invert $\mathbf{Z}'\mathbf{Z}$ directly and the method breaks down. The generalized inverse can be used instead of the usual matrix inverse for slightly underdetermined system but for significantly underdetermined system, we need to regularize using the l_1-norm penalty.

1.2 Voxel-Based Morphometry

GLM has often been use in *voxel-based morphometry* (VBM). Let us review basic VBM that is needed to understand how GLM is used in VBM. VBM involves a voxel-wise comparison of the local concentration of gray or white matters between populations [8]. It requires spatially normalizing images from all the subjects in the study to a template. This is followed by segmenting the gray and white matters and cerebrospinal fluid (CSF) from the spatially normalized images and smoothing out the segmented images. The binary segmented images are refereed as tissue densities (Figure 1.1). Statistical inference is subsequently done at each voxel level on the tissue densities while accounting for multiple comparisons.

The tissue segmentation is based on a Gaussian mixture model that assumes the image intensity values to follow the mixture of three independent

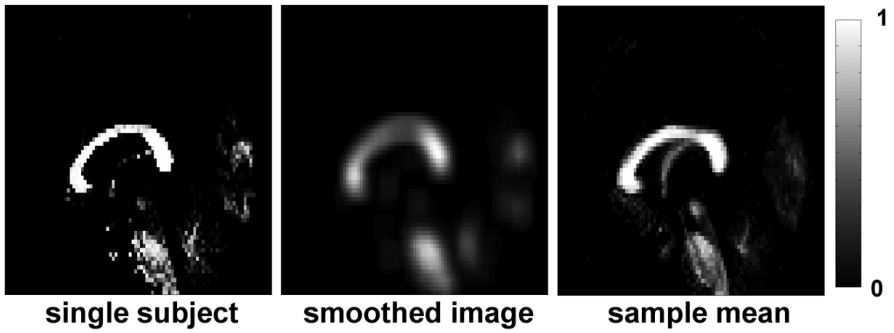

FIGURE 1.1
(a) Original white matter density in the corpus callosum obtained from the SPM package. (b) Gaussian kernel smoothing on the white matter density map. (c) The sample mean of smoothed images of all subjects.

Gaussians and the unknown parameters of Gaussian distributions are estimated by maximizing the likelihood function using the expectation maximization (EM) algorithm. Figure 1.2 shows an example of two component Gaussian mixtures. The widely used Statistical Parametric Mapping (SPM) package (Wellcome Department of Cognitive Neurology, London, UK. `www.fil.ion.ucl.ac.uk/spm`) is based on a Bayesian formulation of the Gaussian mixture model with a prior probability image obtained by averaging already segmented large number of brain images [10, 8]. Based on the prior probability of each voxel belong to a specific tissue type, the Bayesian framework is used to get the posterior probability. This Bayesian update of the probability is iterated many times until the probability converges. The resulting probability map is interpreted as the tissue density. This is not physical density so it should be interpreted probabilistically.

The Bayesian segmentation framework utilizes the Bayes theorem in estimating the poster probability of a voxel belong to a particular tissue type from a given prior probability. Let C be the event of a voxel belong to a particular class. We may assume there are three classes corresponding to gray, white matters and CSF. The prior probability $P(C)$ is obtained by averaging a large sample of normalized binary segmentation and dividing the average by the total number of sample. Let T be the event that a voxel has a particular image intensity value. This is that we usually observe in T_1-weighted MRI. We wish to obtain the *conditional probability* $P(C|T)$ of the voxel belong to the class C given that we have observed T:

$$P(C|T) = \frac{P(C \cap T)}{P(T)}. \tag{1.7}$$

$P(C|T)$ is interpreted as the probability of the voxel belong to a specific

class when the voxel has a particular intensity value. This is what we likely to determine in Bayesian segmentation and it is termed as tissue density in VBM. The numerator can be written as $P(C \cap T) = P(T|C)P(C)$ while, from the law of total probability, the total probability $P(T)$ can be decomposed as

$$P(T) = \sum_C P(T \cap C) = \sum_C P(T|C)P(C).$$

The conditional probability (1.7) can be written in terms of the prior probability as

$$P(C|T) = \frac{P(T|C)P(C)}{\sum_C P(T|C)P(C)}. \tag{1.8}$$

The likelihood term $P(T|C)$ is interpreted as the probability of a voxel obtaining a particular intensity value given the voxel belong to a particular tissue type, and it can be estimated from mixture models. The likelihood term is given by evaluating the probability density for the class C at each voxel intensity value [8].

1.2.1 Mixture Models

To estimate the likelihood term, it is necessary to introduce mixture models and the expectation-maximization algorithm. Mixture models have been widely used for segmenting brain images. The image intensity value at a given voxel can come from different tissue classes with specific proportions p_j. We will assume $0 < p_j < 1$ and $\sum_j p_j = 1$. We may assume that image intensity values for each class to follow a certain distribution f_j. This is the likelihood term $P(T|C)$. Then the k-component mixture model on image intensity values assume image intensity values Y to come from k different distributions $f_1, \cdots, f_k$ with proportions $p_1, \cdots, p_k$. This can be modeled by conditioning on a multinomial distribution. Another way of saying this is that the the k-component mixture model can be obtained by mixing samples obtained from distributions f_j with p_j proportions.

Let X_j be an indicator variable for the j-th class such that $P(X_j = 1) = p_j$ and $P(X_j = 0) = 1 - p_j$. X_j is a Bernoulli random variable. The collection of variables $X = (X_1, \cdots, X_k)$ form a multinomial distribution with parameters $(p_1, \cdots, p_k)$ if we have the additional constraint $X_1 + \cdots + X_k) = 1$. The probability mass function of X is given by

$$f(x_1, \cdots, x_k) = P(X_1 = x_1, \cdots, X_k = x_k) = p_1^{x_1} \cdots p_k^{x_k}.$$

Now we define a random variable Y conditionally on the event $X_j = 1$ such that $Y \sim f_j$ if $X_j = 1$. The conditional density $f(y|x_j = 1) = f_j$ is the distribution for the j-th class. The joint density between X_j and Y is then given by $f(x_j = 1, y) = p_j f_j(y)$, which can be compactly written as

$$f(x, y) = [p_1 f_1(y)]^{x_1} \cdots [p_k f_k(y)]^{x_k}$$

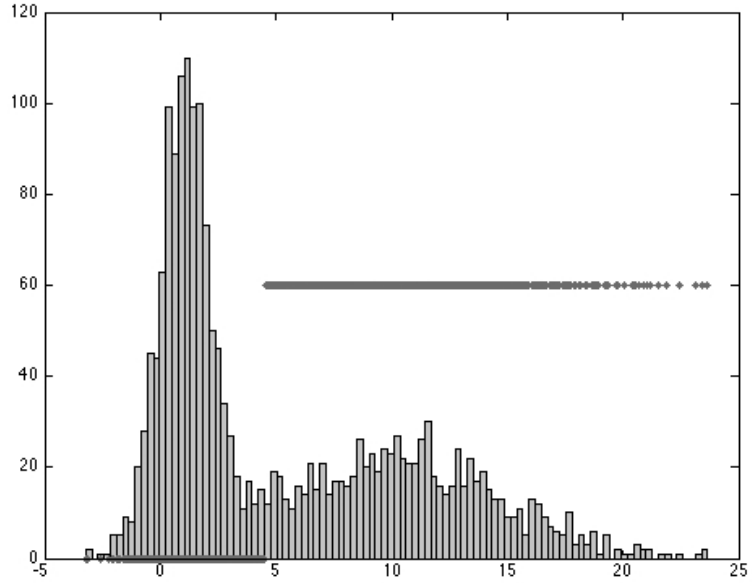

FIGURE 1.2
Simulated two component Gaussian mixture with parameters $\mu_1 = 10, \sigma_1 = \sqrt{5}, \mu_2 = 1, \sigma_2 = 1,_1 = 0.5,_2 = 0.5$. The parameters are then estimated using the EM-algorithm. Using the estimated model, we cluster data depending on $f_1(x) < f_2(x)$ or $f_1(x) \geq f_2(x)$.

for all j. The marginal density of Y is subsequently given as

$$f(y) = \sum_x f(x, y) = \sum_{i=1}^{k} p_i f_i(y). \tag{1.9}$$

The unknown parameters in (1.9) will be denoted as Θ. The unknown parameters include the mixing proportions p_j as well as parameters of the distribution f_i. Then we write the *k-component mixture model* as

$$f(y|\Theta) = \sum_{i=1}^{k} p_i f_i(y). \tag{1.10}$$

to indicate the dependence of the model on the parameters Θ. The most widely used technique for estimating Θ in (1.10) is the *maximum likelihood estimation* (MLE). Suppose we have a sample $Y = \{Y_1, \cdots, Y_n\}$ drawn from the distribution $f(y|\Theta)$. The likelihood estimation of Θ is given by maximizing

the loglikelihood:

$$\widehat{\Theta} = \arg\max_{\Theta} \prod_{i=1}^{n} f(y_i|\Theta) = \arg\max_{\Theta} \sum_{i=1}^{n} \ln f(y_i|\Theta).$$

For most mixture models, the optimization cannot be done analytically and it requires an iterative approximation technique called the expectation maximization (EM) algorithm.

1.2.2 EM-Algorithm

The expectation maximization (EM) algorithm was first introduced by [41]. For the introductory overview on the algorithm, see [103] and [47]. The EM-algorithm proceeds as follows.

Following the argument in [103], we augment the observed data Y with latent (unobserved or missing) data Y^m such that the complete data $Y^c = (Y, Y^m)$. The latent data is introduced as an artifice to make the problem tractable. The probability density of the complete data Y^c is denoted as $f(y^c) = f(y, y^m)$. The conditional density for the latent data Y^m, condition on observation Y, is

$$f(y^m|y, \Theta) = \frac{f(y, y^m|\Theta)}{f(y|\Theta)}.$$

Again we introduced Θ to indicate the dependence of the probability on the parameters. Taking the logarithm on both sides, we get the loglikelihood for the observed data

$$\ln f(Y|\Theta) = \ln f(Y^c|\Theta) - \ln f(Y^m|Y, \Theta).$$

Since the logarithm is a strictly increasing function, the value that maximizes $f(Y|\Theta)$ also maximizes $\ln f(Y|\Theta)$. Now taking the expectation with respect to $f(y^m|y, \Theta_0)$ for some fixed Θ_0 on the both sides, we have

$$\mathbb{E}[\ln f(Y|\Theta)|Y, \Theta_0] = \mathbb{E}[\ln f(Y^c|\Theta)|Y, \Theta_0] - \mathbb{E}[\ln f(Y^m|Y, \Theta)|Y, \Theta_0]. \quad (1.11)$$

Now denote the expected loglikelihood for the complete data as

$$Q(\Theta|\Theta_0, Y) = \mathbb{E}[\ln f(Y^c|\Theta)|Y, \Theta_0].$$

We maximize the likelihood in iterative two-steps:

(1) E-step: compute the expectation $Q(\Theta|\widehat{\Theta}_{j-1}, Y)$.

(2) M-step: maximize $Q(\Theta|\widehat{\Theta}_{j-1}, Y)$ and take

$$\widehat{\Theta}_j = \arg\max_{\Theta} Q(\Theta|\widehat{\Theta}_{j-1}, Y). \quad (1.12)$$

Starting with the initial estimate $\widehat{\Theta}_0$, we have a sequence of estimators $\widehat{\Theta}_1, \widehat{\Theta}_2, \cdots$ and it can be shown to converges to the true MLE $\widehat{\Theta}$. However, the proof is beyond the scope of the book and we will only show that the Q function monotonically increases. The argument is as follows. By the definition (1.12), we have

$$Q(\widehat{\Theta}_j|\widehat{\Theta}_j, y) \leq Q(\widehat{\Theta}_{j+1}|\widehat{\Theta}_j, y).$$

Now let $R(\Theta|\Theta_0, Y) = \mathbb{E}[\ln f(Y^m|Y, \Theta)|Y, \Theta_0]$. This is the second term in (1.11). From the Jensen's inequality, we can show that

$$R(\Theta|\Theta_0, y) - R(\Theta_0|\Theta_0, y) \leq \ln \int \frac{f(y^m|y, \Theta)}{f(y^m|y, \Theta_0)} f(y^m|y, \Theta_0)\, dy^m = 0.$$

Hence we have $R(\widehat{\Theta}_{j+1}|\widehat{\Theta}_j, y) \leq R(\widehat{\Theta}_j|\widehat{\Theta}_j, y)$. Consequently

$$\begin{aligned} \ln f(y|\widehat{\Theta}_j) &= Q(\widehat{\Theta}_j|\widehat{\Theta}_j, y) - R(\widehat{\Theta}_j|\widehat{\Theta}_j, y) \\ &\leq Q(\widehat{\Theta}_{j+1}|\widehat{\Theta}_j, y) - R(\widehat{\Theta}_{j+1}|\widehat{\Theta}_j, y) \\ &\leq \ln f(y|\widehat{\Theta}_{j+1}). \end{aligned}$$

The inequality guarantees the the sequence of estimators $\widehat{\Theta}_j$ monotonically increases the likelihood function. Further, since the monotonically increasing sequence is bounded, i.e. $\ln f(y|\widehat{\Theta}_j) \leq \ln f(y|\widehat{\Theta})$, where $\widehat{\Theta}$ is the MLE, the sequence must be converging to a constant, but it is not clear if the limit is in fact $\ln f(y|\widehat{\Theta})$. To guarantee that the limit converges to the true maximum likelihood estimator, additional conditions are needed boyles.1983, wu.1983.

The difficulty of implementing the EM-algorithm is at the E-step where we need to compute the conditional expectation $Q(\Theta|\widehat{\Theta}_{j-1}, y)$. The Monte Carlo version of the EM algorithm overcome this problem by simulating the missing data Y^m from the conditional density $f(y^m|y, \Theta)$ so that

$$\widehat{Q}(\Theta|\Theta_0, y) = \frac{1}{k} \sum_{j=1}^{k} \ln f(Y, Y^m|\Theta).$$

As an illustration, two component Gaussian mixture model is shown in detail but the three component mixture model is similar. The image intensity will be modelled as a Gaussian mixture of the form

$$f(y) = p_1 f_1(y) + p_2 f_2(y)$$

where $p_1 + p_2 = 1$ and $f_1 \sim N(\mu_1, \sigma_1^2)$ and $f_2 \sim N(\mu_2, \sigma_2^2)$ are all known. There are 5 unknown parameters $\Theta = \{p_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\}$ to be estimated. Once p_1 is estimated, p_2 is automatically given as $1 - p_1$. The likelihood function is given by

$$f(\Theta|y) = \prod_{i=1}^{n} \Big[p_1 f_1(y_i) + p_2 f_2(y_i) \Big].$$

The loglikelihood is

$$L(\Theta|y) = \sum_{i=1}^{n} \ln \Big[p_1 f_1(y_i) + p_2 f_2(y_i) \Big].$$

The loglikelihood is maximized by solving

$$\frac{\partial L(\Theta|y)}{\partial p_i} = 0, \ \frac{\partial L(\Theta|y)}{\partial \mu_i} = 0, \frac{\partial L(\Theta|y)}{\partial \sigma_i^2} = 0$$

but this is not tractable. So we argument the data with the latent data and apply the EM-algorithm.

Let X be a Bernoulli random variable with $P(X = 1) = p$ and $P(X = 0) = q = 1 - p$. This choice of latent random variable makes the subsequent EM-aglorithm to be tractable. Now define the conditional distribution $Y \sim f_1$ if $X = 1$ and $Y \sim f_2$ if $X = 0$. This defines the conditional density $f(y|x)$. The joint density $f(x, y)$ is $f(1, y) = pf_1(y)$ and $f(0, y) = qf_2(y)$. This can be compactly written as

$$f(x, y) = \Big[pf_1(y) \Big]^x \Big[qf_2(y) \Big]^{1-x}.$$

The marginal density of Y is obviously

$$f(y) = \sum_{x=0,1} f(x, y) = pf_1(y) + qf_2(y).$$

The conditional density of X given Y is then

$$f(x|y) = \frac{[pf_1(y)]^x [qf_2(y)]^{1-x}}{pf_1(y) + qf_2(y)}.$$

The conditional expectation of X with respect to $f(x|y)$ is then

$$\mathbb{E}(X|y, p) = \frac{pf_1(y)}{pf_1(y) + qf_2(y)}. \tag{1.13}$$

The likelihood for the complete data (x, y) is

$$f(\Theta|x, y) = \prod_{i=1}^{n} \big[pf_1(y_i) \big]^{x_i} \big[qf_2(y_i) \big]^{1-x_i}$$

and the corresponding loglikelihood is given by

$$L(\Theta|x, y) = \sum_{i=1}^{n} x_i \ln \Big[\frac{pf_1(y_i)}{qf_2(y_i)} \Big] + \ln \big[qf_2(y_i) \big].$$

Take the expectation with respect to the latent variable X to get the Q-function

$$\begin{aligned} Q(\Theta|\Theta_0, y) &= \mathbb{E}\big[\ln L(\Theta|X, Y)\big|y, \Theta_0\big] \\ &= \sum_{i=1}^{n} \mathbb{E}(X_i|y, \Theta_0) \ln\Big[\frac{pf_1(y_i)}{qf_2(y_i)}\Big] + \ln\big[qf_2(y_i)\big]. \end{aligned} \quad (1.14)$$

From (1.13), we have

$$\mathbb{E}(X_i|y, \Theta_0) = \frac{p_0 f_1(y_i)}{p_0 f_1(y_i) + q_0 f_2(y_i)} = \pi_{1i}$$

is the posterior probability of the i-th observation coming from the first class. Hence the expression (1.14) can be written as

$$Q(\Theta|\Theta_0, y) = \sum_{i=1}^{n} \pi_i \ln\Big[\frac{pf_1(y_i)}{qf_2(y_i)}\Big] + \ln\big[qf_2(y_i)\big]. \quad (1.15)$$

Maximizing Q with respect to p by solving $\partial Q/\partial p = 0$, we obtain

$$p = \frac{1}{n}\sum_{i=1}^{n} \pi_{1i}. \quad (1.16)$$

(1.16) states that the prior probability for the 1st class is estimated as the average of the posterior probabilities in the 1st class. Note that we did not use the explicit forms for f_1 and f_2 so this result is general for any type of mixture distributions. Based on (1.16), we set up the iteration

$$\widehat{p}_{j+1} = \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{p}_j f_1(y_i)}{\widehat{p}_j f_1(y_i) + (1-\widehat{p}_j) f_2(y_i)}$$

with any arbitrary initial $\widehat{p}_0 \in (0, 1)$. For another parameters, we obtain similar iterative formulas:

$$\mu_j = \frac{\sum_i \pi_{ji} y_i}{\sum_i \pi_{ji}}$$

and

$$\sigma_j^2 = \frac{\sum_i \pi_{ji}(y_i - \mu_j)^2}{\sum_i \pi_{ji}}.$$

1.3 Application to Autism

Autism is a neurodevelopmental disorder of brain function that has begun to attract *in vivo* structural magnetic resonance imaging (MRI) studies in the

region of the corpus callosum [43, 62, 83, 93, 94]. The corpus callosum is a white matter structure that can be used as an index of neural connectivity between brain regions [62] (Figure 1.3). There is little understanding about the link between the functional deficit and the underlying abnormal anatomy in autism, which provides motivation for our study. These studies use the Witelson partition or a similar partition scheme of the corpus callosum [125]. Witelson partitioned the midsagittal cross-sectional images of the corpus callosum along the maximum anterior-posterior line [112] and defined the region of the genu, rostrum, midbodies, isthmus and splenium from the anterior to posterior direction. Based on the Witelson partition, there has been a consistent finding in abnormal reduction in anterior, midbody and posterior of the corpus callosum [17].

Piven *et al.* (1997) compared 35 autistic individuals with 36 normal control subjects controlling for total brain volume, gender and IQ and detected a statistically significant smaller midbody and posterior regions of the corpus callosum in the autistic group [94]. Manes *et al.* (1999) compared 27 low functioning autistic individuals with 17 normal controls adjusting for the total brain volume [83]. They found a smaller corpus callosum compared to the control group in genu, rostrum, anterior midbody, posterior midbody and isthmus but did not find statistically significant differences in the rostrum and the splenium although the sample mean of the rostrum and splenium size are smaller than that of the control group. Hardan *et al.* (2000) compared 22 high functioning autistic to 22 individually matched control subjects and showed smaller genu and rostrum of the corpus callosum adjusting for the total brain volume based on the Witelson partition [62]. The smaller corpus callosum size was considered as an indication of a decrease in interhemispheric connectivity. They did not detect other regions of significant size difference. For an extensive review of structural MRI studies for autism that have been published between 1966 and 2003, one may refer to Brambilla *et al.* (2003) [17].

The shortcoming of the Witelson partition is the artificial partitioning. The Witelson partition may dilute the power of detection if the anatomical difference occurs near the partition boundary. Alternative voxel-wise approaches that avoid predefined regions of interests (ROI) have begun to be used in structural autism studies. Vidal *et al.* (2003) used the tensor-based morphometry (TBM) to show reduced callosal thickness in the genu, midbody and splenium in autistic children [121]. Hoffmann *et al.* (2004) used a similar TBM to show curvature difference in the midbody [65]. Abell *et al.* (1999) used voxel-based morphometry (VBM) in high functioning autism to show decreased gray matter volume in the right paracingulate sulcus, the left occipito-temporal cortex and increased amygdala and periamygdaloid cortex.

The advantage of the VBM framework over the Witelson partition approach is that it is completely automated and does not require artificial partitioning of the corpus callosum that introduces undesirable bias. Further it is not restricted to *a priori* ROIs enabling us to perform the statistical analysis

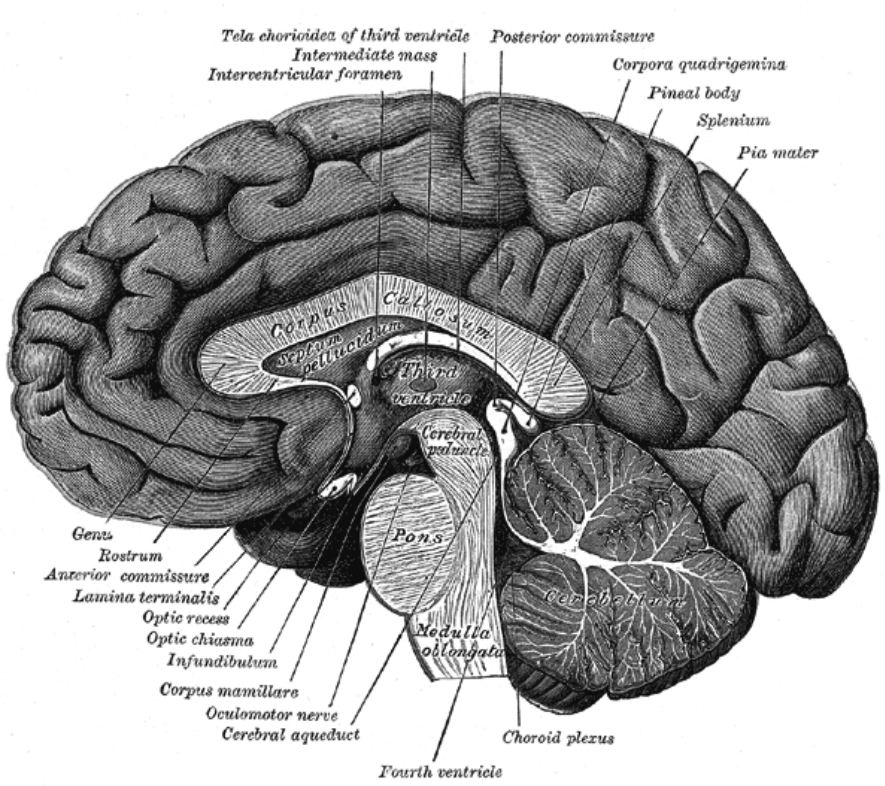

FIGURE 1.3
Midsagittal cross section of brain. Corpus callosum is the a colluction of neural fibers that connects the left and the right cerebral hemispheres. The posterior part of the corpus callosum is called the splenium; the anterior part is called the genu. Between them is the midbody. The illustration is from Gray's Anatomy [58].

at each voxel level and to pinpoint the exact location of the anatomical differences within ROI even if there is no ROI size differences. Although VBM was originally developed for whole brain 3D morphometry, our study concentrates on the midsagittal cross sectional corpus callosum regions to be able to compare the result with the previous 2D Witelson partition studies [62, 83, 94].

1.3.1 White matter Density Maps

Gender and handedness affect the corpus callosum anatomy [124, 125] so all the subjects used in the data set are right-handed males. Sixteen autistic subjects were recruited for this study from a list of individuals with a diagnosis of high functioning autism in the Madison and Milwaukee area maintained

for research purposes by the Waisman center at the University of Wisconsin-Madison. Diagnoses were confirmed with the Autism Diagnostic Interview - Revised (ADI-R) or clinical interview administered by a trained and certified psychologist at the Waisman center. All participants met DSM-IV criteria for autism or Asperger's pervasive developmental disorder. Twelve healthy, typically developing males with no current or past psychological diagnoses served as a control group. The average age for the control subject is 17.1 ± 2.8 and the autistic subjects is 16.1 ± 4.5 which is in compatible age range.

High resolution anatomical MRI scans were obtained using a 3-Tesla GE SIGNA (General Electric Medical Systems, Waukesha, WI) scanner with a quadrature head RF coil. A three-dimensional, spoiled gradient-echo (SPGR) pulse sequence was used to generate T1-weighted images. The imaging parameters were TR/TE 21/8 ms, flip angle 30°, 240 mm field of view, 256x192 in-plane acquisition matrix (interpolated on the scanner to 256x256), and 128 axial slices (1.2 mm thick) covering the whole brain. Then the midsagittal cross-sections of the white matter are segmented using the SPM-package (Figure 1.4).

1.3.2 Manipulating Density Maps

The segmented imaging data is stored in the directory `CCdensity` as text files with the file extension `*.txt`. For instance, `CCautism12.txt` is the the white matter density for the 12th autistic subject while `CCcontrol03.txt` is the white matter density for the 3rd control subject. This is the data set published in [30] where white matter density in the corpus callosum was analyzed. Figure 1.4 shows the images of white matter density at corpus callosum.

Sequently we run the following codes. Subject identifies are usually given in numbers:

```
c1=[01 02 05 08 09]';
c2=[10 11 12 13 14 15 16]';

a1=[01 02 03 04 05 06 07 08 09]'
a2=[10 11 13 14 16 17 18]';
```

The file names are then given with subject identifiers as strings:

```
file_c1=strcat('CCcontrol0',int2str(c1),'.txt');
file_c2=strcat('CCcontrol',int2str(c2),'.txt');
file_c=[file_c1; file_c2]

file_a1=strcat('CCautism0',int2str(a1),'.txt');
file_a2=strcat('CCautism',int2str(a2),'.txt');
file_a=[file_a1; file_a2]
```

This produces the list of file names:

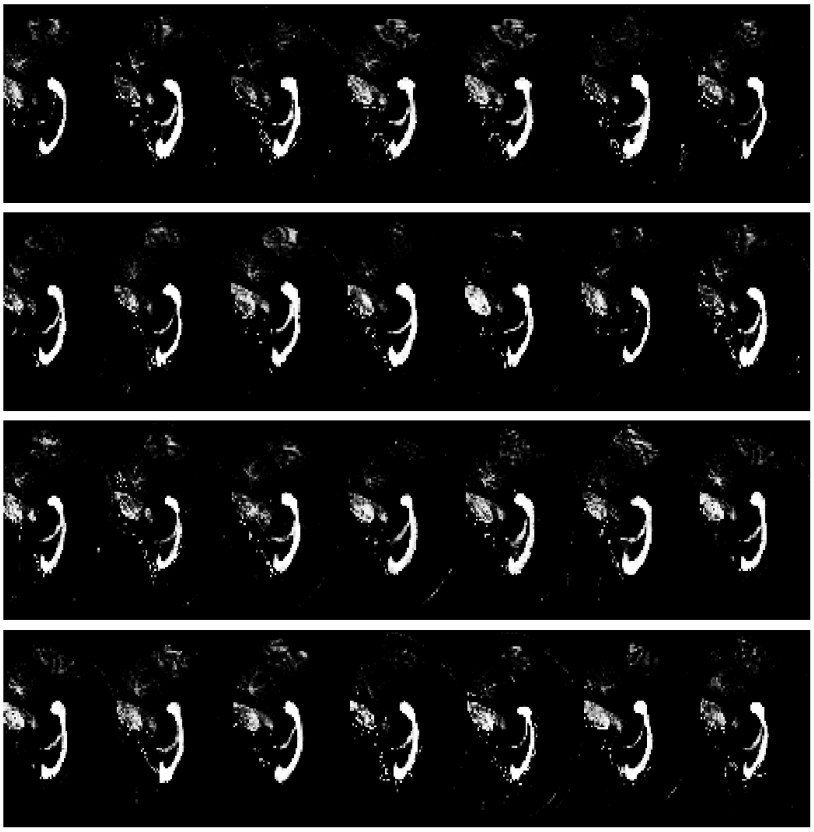

FIGURE 1.4
Midsaggital cross-section images of corpus callosum for 28 subjects. The first 12 subjects are controls while the next 16 subjects are autistic subjects. We are interested in localizing the density difference between the two groups.

```
file_c =

CCcontrol01.txt
CCcontrol02.txt
CCcontrol05.txt
.
.
.
```

We need to concatenate in two different ways since `int2str(08)` does not produce `08` but `8`. The total number of subjects in each group is

```
n_c=size(file_c,1)
```

```
n_a=size(file_a,1)

n_c =
    12

n_a =
    16
```

The dimension of images is 95×68. So we let

```
d1=95; d2=68;
```

Once we have the list of files, we load them sequentially into matrices:

```
density_a=zeros(n_c,d1,d2);
density_a=zeros(n_a,d1,d2);

for i=1:n_c
    temp=load(file_c(i,:));
    temp=reshape(temp,d1,d2);
    density_c(i,:,:)=temp;
end;

for i=1:n_a
    temp=load(file_a(i,:));
    temp=reshape(temp,d1,d2);
    density_a(i,:,:)=temp;
end;
```

All 28 subject images can be visualized using

```
img=[];
for i=1:n_c
    img = [img, squeeze(density_c(i,:,:))];
end;
figure; imagesc(img); colormap('bone'); colorbar;

img=[];
for i=1:n_a
    img = [img, squeeze(density_a(i,:,:))];
end;
figure; imagesc(img); colormap('bone'); colorbar
```

The resulting density maps are shown in Figure 1.4. The first 12 images are controls and the next 16 images are autistic subjects. The additional command `set(gcf,'Color','w')` will set the background of an image white. To save the image as a file, `print('-dtiff', '-r300', 'CC')` can be used.

1.3.3 MATLAB Implementation

A simple example of GLM is the usual two-sample t-test setting. Given two groups, we are interested in testing the significance of group difference on tissue density. So we consider the following GLM:

$$\texttt{density}_i = \lambda_1 + \beta_1 \cdot \texttt{group}_i + \epsilon, \tag{1.17}$$

where the dummy variable `group` is 1 for autism and 0 for control. This is the case for $k = 1$, $z_{i1} = 1$ and $p = 1$. Another more complicated example is the case of liner regression for two groups, which can be combined into a single GLM:

$$\texttt{density}_i = \lambda_1 + \lambda_2 \cdot \texttt{age}_i + \beta_1 \cdot \texttt{group}_i \tag{1.18}$$

This is the case for $k = 2$ and $p = 1$ (Figure 1.7). We will implement (1.18) in MATLAB and estimate the parameters using the least squares method.

Once we loaded images, we set up a general linear model and estimate the parameters in a least squares fashion. The combined age information of both groups is stored in `age` which is a column vector of size 28×1. The design matrix `X` consists of a column of ones and vector `age`.

```
age_c = [15 18 18 16 15 13 18 15 21 17 16 23]'
age_a = [15 20 17 13 12 15 25 14 15 14 24 18 10 12 22 12]'
age=[age_c;age_a];

const=ones(n_c+n_a,1);
X=[const, age];
```

The density map for controls `density_c` is of size $12 \times 95 \times 68$. We reshape it into a 2-dimensional matrix of size $12 \times (95 \cdot 68)$.

```
p_c=reshape(density_c,12,d1*d2);
p_a=reshape(density_a,16,d1*d2);
p=[p_c; p_a];
```

Then we estimate the parameters in the linear model

$$\texttt{density} = \texttt{lambda0} + \texttt{lambda1} \cdot \texttt{age} \tag{1.19}$$

using the pseudoinverse operation `pinv` (Figure 1.5). The estimated parameter maps are given in Figure 1.5.

```
lambda=zeros(2,d1*d2);
for i=1:(d1*d2)
    lambda(:,i)=pinv(X)*p(:,i);
end
```

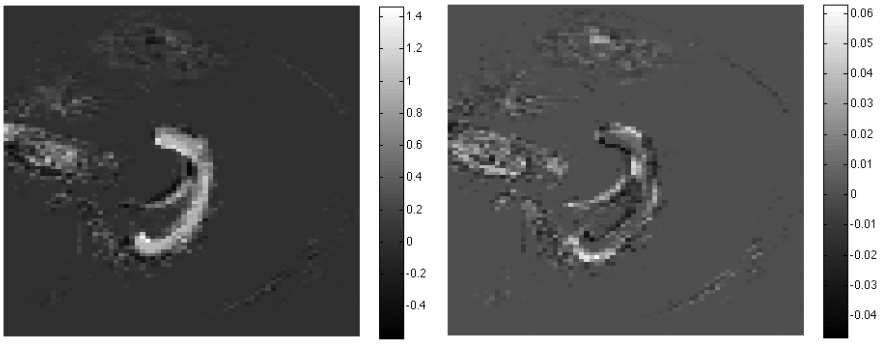

FIGURE 1.5
The least squares estimation of the parameters `lambda0` and `lambda1`.

```
lambda0=reshape(lambda(1,:),d1,d2);
lambda1=reshape(lambda(2,:),d1,d2);

figure;imagesc(lambda0);colorbar; colormap('bone')
figure;imagesc(lambda1);colorbar; colormap('bone')
```

The sum of the residual at each pixel is given by

```
SSE0 = sum((p - (X*lambda)).^2);
SSE0=reshape(SSE0,d1,d2);
```

We add the additional term `group` in (1.19):

$$\texttt{density} = \texttt{lambda0} + \texttt{lambda1} \cdot \texttt{age} + \texttt{lambda2} \cdot \texttt{group}.$$

For this, we have another column in the design matrix `X`:

```
group = [zeros(12,1); ones(16,1)]
X=[const, age, group];
```

The design matrix now looks like

```
...
1    16     0
1    23     0
1    15     1
1    20     1
...
```

The parameters are again estimated by the pseudoinverse operation:

```
lambda=zeros(3, d1*d2);
for i=1:(d1*d2)
    lambda(:,i) = pinv(X)*p(:,i);
end
```

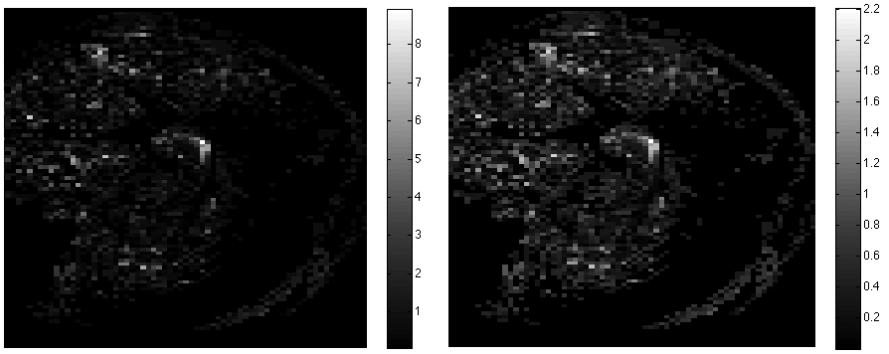

FIGURE 1.6
The F-statistic of testing the significance of the term `group` and the corresponding p-value in $\log_{10}$ scale. The signal is detected in the genu of the corpus callosum. Since we didn't smooth the density maps, we are getting a lot of noise.

The total sum of squared error is then

```
SSE1 = sum((p - (X*lambda)).^2);
SSE1=reshape(SSE1,d1,d2);
```

The statistic of testing the significance of `group` is based on the ratio of the sum of squared errors between `SSE0` and `SSE1` (Figure 1.6).

```
F=25*(SSE0-SSE1)./SSE0;
imagesc(F);colorbar; colormap('bone')
pvalue=1 - fcdf(F,1,25);
figure; imagesc(-log10(pvalue)); colormap('bone')
```

1.4 Modeling Growth Rate

Let us cover few other issues that are left out. As shown in Figures 1.8 and 1.7, each group has different growth rate, the change of white matter density over age. The question is if the growth rate difference is statistically significant. This can be tested within the GLM framework. Note the GLM incorporates two-way ANOVA and ANCOVA. Assume there are total n subjects. Let $\texttt{density}_i$ be the white matter density for the i-th subject. Let $\texttt{age}_i$ be the age of the i-th subject. Let $\texttt{group}_i$ be the group variable of the i-th subject taking value 1 for autistic and 0 for controls. Then we have the following GLM:

$$\texttt{density}_i = \beta_0 + \beta_1 \texttt{age}_i + \beta_2 \texttt{group}_i + \beta_3 \texttt{age}_i \cdot \texttt{group}_i + \epsilon_i. \qquad (1.20)$$

From (1.20), we have two separate models for each group. For controls ($\texttt{group}_i = 0$),

$$\texttt{density}_i = \beta_0 + \beta_1 \texttt{age}_i + \epsilon_i.$$

For autistic subjects ($\texttt{group}_i = 1$), we have

$$\texttt{density}_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)\texttt{age}_i + \epsilon_i.$$

Testing the equality of the growth rate is equivalent to testing

$$H_0 : \beta_1 = \beta_1 + \beta_3$$

or equivalently

$$H_0 : \beta_3 = 0.$$

We need to test for the significance of the interaction term β_3 in the model.

The fit of model is measured by the sum of squared errors (SSE). Let $\texttt{SSE}_0$ and $\texttt{SSE}_1$ be the SSE for the reduced (when $\beta_3 = 0$)and the full model (when $\beta_3 \neq 0$) respectively. Then

$$\texttt{SSE}_0 = \sum_{i=1}^{m+n} (\texttt{density}_i - \widehat{\beta}_0 + \widehat{\beta}_1 \texttt{age}_i - \widehat{\beta}_2 \texttt{group}_i)^2,$$

where $\widehat{\beta}_i$ are the estimated regression coefficients from the reduced model. Similarly for the full model,

$$\texttt{SSE}_1 = \sum_{i=1}^{m+n} (\texttt{density}_i - \widehat{\gamma}_0 + \widehat{\gamma}_1 \texttt{age}_i - \widehat{\gamma}_2 \texttt{group}_i - \widehat{\gamma}_3 \texttt{age}_i \cdot \texttt{group}_i)^2,$$

where $\widehat{\gamma}_i$ are the estimated regression coefficients from the full model. Then the F-statistic is given by the ratio of SSE:

$$F = \frac{(\texttt{SSE}_0 - \texttt{SSE}_1)/1}{\texttt{SSE}_0/(n-1-3)} \sim F_{1,n-1-3}.$$

1.5 Gaussian Kernel Smoothing

All brain images are inherently noisy due to errors associated with image acquisition. Compounding the image acquisition errors, there are errors caused by image registration and segmentation. So it is necessary to smooth out the segmented images before any morphometric analysis is performed. Among many possible image smoothing methods [74, 91], *Gaussian kernel smoothing*

has emerged as a de facto smoothing technique in brain imaging. Consider a n-dimensional Gaussian kernel

$$K(x) = \frac{1}{(2\pi)^{n/2}} \exp\Big(- \frac{\|x\|^2}{2}\Big),$$

where $\|\cdot\|$ is the Euclidean norm of $x \in \mathbb{R}^n$. The rescaled kernel K_σ is defined as

$$K_\sigma(x) = \frac{1}{\sigma^n} K\big(\frac{x}{\sigma}\big). \tag{1.21}$$

Then an integral version of Gaussian kernel smoothing in n-dimension is defined as

$$F(x, \sigma) = \int_{\mathbb{R}^n} K_\sigma(x - y) f(y)\, dy,$$

where $F(x, \sigma)$ is the scale-space representation of image $f(x)$ first introduced in [126]. Each $F(x, \sigma)$ for different values of σ produces a blurred copy of its original. The resulting scale-space representation from coarse to fine resolution can be used in multiscale approaches such as hierarchical searches and image segmentation. See [79], [96], [97], [108], and [132] for the review of the major problems in scale-space and multiscale descriptions of images.

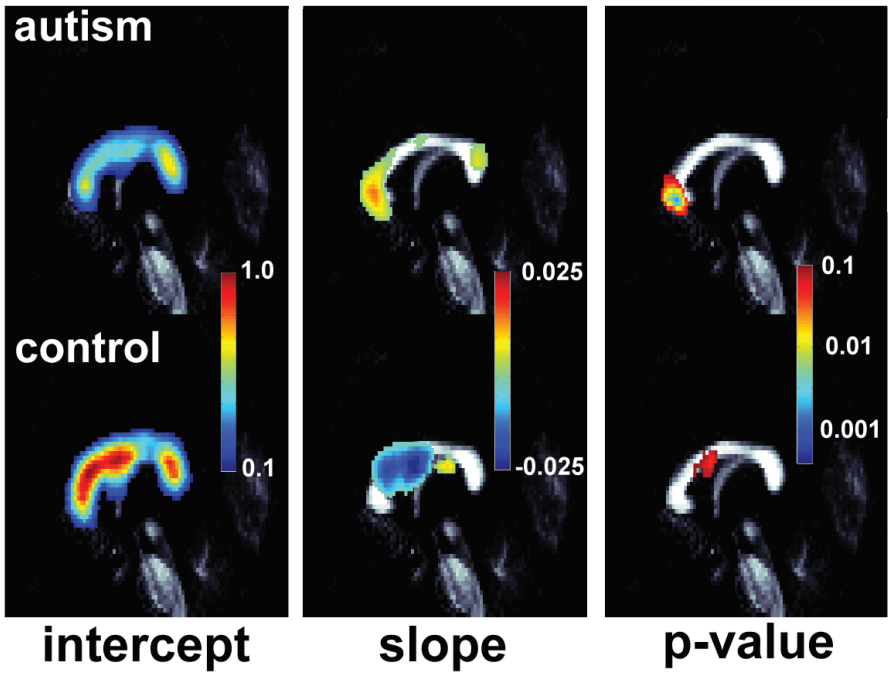

FIGURE 1.7
GLM fit of white matter density over age and group at each voxel: `density` = $\lambda_1 + \lambda_2 \cdot$ `age` $+ \lambda_3 \cdot$ `group`. The intercept and slopes of the linear regression is for each group. The autistic group shows lower white matter density compared to the control at lower age but gains white matter over time while the control group shows decreasing white matter density with age [30].

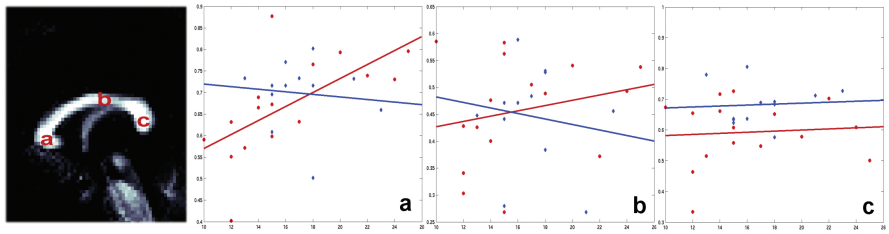

FIGURE 1.8
Linear regression fit for each group (red=autism, blue = control) (a) Genu (b) Midbody and (c) spleninum of the corpus callosum.

2

Random Fields Theory

NOTE: MATLAB COMPONENTS ARE NOT ADDED YET.

In this chapter, we cover *multiple comparisons* issue that has been crucial in determining over all statistical significance in a correlated test statistic over a whole brain image. In practice, F statistics in adjacent voxels are correlated. So there is the problem of multiple comparisons, which we simply neglected in obtaining the p-value map. The multiple comparisons will be covered in detail later but here we will go over the concept briefly. For multiple comparisons that account for spatially correlated error, we can use the result of the random field theory [128, 133], the false discovery rates [14, 15, 52] or permutation tests [90]. If $F(x)$ is a smooth F-field, the *corrected* p-value for adjusting multiple comparisons over the all pixels in the corpus callosum $\mathcal{M}$ is given by

$$P\big(\sup_{x\in\mathcal{M}} Y(x) > h\big) \approx \sum_{d=0}^{2} \mu_d(\Omega)\rho_d(h)$$

where μ_d are the d-dimensional Minkowski functionals and ρ_d are the d-dimentional Euler characteristic (EC) density [133]. For non regular jagged shapes such as the corpus callosum, the Minkowski functionals can be estimated in the following fashion. Treating pixels inside $\mathcal{M}$ as points on a lattice, V is the number of suh points, E is the number of edges connecting each adjacent lattice points, F is the number of faces formed by 4 connected edges. Then $\mu_0 = \chi(\Omega) = V - E + F = 1, \mu_1 = (E - 2F)\delta, \mu_2 = F\delta^2$ where $\delta = 2$ is the resolution of the image [127]. To find the number of edges and pixels contained in $\mathcal{M}$, we start from an initial face F_0 in the splenium of CC and add an additional face in the lattice one by one while counting the additional edges and faces [30]. A similar approach for computing Minkowski functionals for jagged irregular shapes has been implemented in FMRISTAT package (`http://www.math.mcgill.ca/keith/fmristat`).

In many brain imaging studies, it is necessary to model measurements at each voxel as a random field. For instance, in the deformation-based morphometry (DBM), deformation fields are usually modeled as continuous random fields. The generalization of a continuous stochastic process defined in $\mathbb{R}$ to a higher dimensional abstract space is called a *random field*. For the introduction to random fields, see [5], [42] and [135]. In the usual random field

modeling, measurement Y at position $x \in \mathcal{M}$ is modeled as

$$Y(x) = \mu(x) + \epsilon(x)$$

where μ is the unknown signal to be estimated and ϵ is the measurement error. The measurement error at each fixed x can be modeled as a random variable. Then the collection of random variables $\{\epsilon(x) : x \in \mathcal{M}\}$ is called a *stochastic process* or *random field*. The more precise measure-theoretic definition can be found in [5]. Random field modeling can be done beyond the usual Euclidean space to curved cortical and subcortical manifolds [69, 34].

2.1 Covariance Functions

Given a probability space, a random field $T(x)$ defined in $\mathbb{R}^n$ is a function such that for every fixed $x \in \mathbb{R}^n$, $T(x)$ is a random variable on the probability space. The *covariance function* $R(x, y)$ of a random field T is defined as

$$R(x, y) = \mathbb{E}\big[T(x) - \mathbb{E}T(x)\big]\big[T(y) - \mathbb{E}T(y)\big].$$

Consider a random field T. If the joint distribution of $T(x_1), \cdots T(x_m)$ given by

$$F_{x_1,\cdots,x_m}(z_1, \cdots, z_m) = P\big[T(x_1) \leq z_1, \cdots, T(x_m) \leq z_m\big]$$

is invariant under the translation

$$(x_1, \cdots, x_m) \rightarrow (x_1 + \tau, \cdots, x_m + \tau),$$

T is said to be *stationary* or homogeneous. For a stationary random field T, we can show $\mathbb{E}T(x) = \mathbb{E}T(0)$ and $R(x, y) = f(x - y)$ for some function f. Although the converse is not always true, such a case is not often encountered in practical applications [135] so we may equate the stationarity with the condition

$$\mathbb{E}T(x) = \mathbb{E}T(0),\ R(x, y) = f(x - y).$$

A special case of stationary fields is an isotropic field which requires the covariance function to be rotation invariant, i.e.

$$R(x, y) = f(\|x - y\|)$$

for some function f. $\|\cdot\|$ is the geodesic distance in the underlying manifold.

Gaussian Fields. An important class of random fields is Gaussian fields. A random vector $T = (T_1, \cdots, T_m)$ is multivariate normal if $\sum_{i=1}^m c_i T_i$ is Gaussian for every possible choice of c_i. Similarly, a random fields T is a *Gaussian random field* if $T(x_1), \cdots, T(x_m)$ are multivariate normal for every

$(x_1, \cdots, x_m) \in \mathbb{R}^m$. An equivalent definition is as follows. T is a Gaussian random field if the finite joint distribution $F_{x_1,\cdots,x_m}(z_1, \cdots, z_m)$ is a multivariate normal for every $(x_1, \cdots, x_m)$. T is a mean zero Gaussian field if $\mathbb{E}T(x) = 0$ for all x. Because any mean zero multivariate normal distribution can be completely characterized by its covariance matrix, a mean zero Gaussian random field T can be similarly determined by its covariance function R. Two fields T and S are independent if $T(x)$ and $S(y)$ are independent for every x and y. For mean zero Gaussian fields T and S, they are independent if and only if the cross-covariance function

$$R(x, y) = \mathbb{E}\big[T(x)T(y)\big]$$

vanishes for all x and y.

The Gaussian white noise is a Gaussian random field with the Dirac-delta function δ as the covariance function. Note the Dirac delta function is defined as $(x) = \infty, x = 0$, $(x) = 0x \neq 0$ and $\int \delta(x) = 1$. Numerically we can simulate the Dirac delta function as the limit of the sequence of Gaussian kernel K_σ when $\sigma \to \infty$. The Gaussian white noise is simulated as independent and identical Gaussian random variable at each voxel.

Derivative and Integration of Fields. Let $\mathcal{G}$ be a collection of Gaussian random fields. For given $X, Y \in \mathcal{G}$, we have $c_1X + c_2Y \in \mathcal{G}$ again for all c_1 and c_2. Therefore, $\mathcal{G}$ forms an infinite-dimensional vector space. Not only the linear combination of Gaussian fields is again Gaussian but also the derivative and integration of random fields are Gaussian. To see this, we define mean-square convergence. A sequence of random fields T_h, indexed by h converges to T as $h \to 0$ in mean-square if

$$\lim_{h\to 0} \mathbb{E}\big|T_h - T\big|^2 = 0.$$

We will denote the convergence as

$$\lim_{h\to 0} T_h = T.$$

Note that the convergence in mean-square implies the convergence in mean. This can be seen from

$$\mathbb{E}\big|T_h - T\big|^2 = \mathbb{V}\big[T_h - T\big]^2 + \big(\mathbb{E}|T_h - T|\big)^2.$$

Now let $T_h \to T$ in mean square. Each term in the right hand side should also converges to zero proving the statement. Now we define the derivative of field in mean square sense as

$$\frac{dT(x)}{dx} = \lim_{h\to 0} \frac{T(x+h) - T(x)}{h}.$$

Note that if $T(x)$ and $T(x+h)$ are Gaussian random fields, $T(x+h) - T(x)$ is

again Gaussian, and hence the limit on the right hand side is again Gaussian. If R is the covariance function of the mean zero Gaussian field T, the covariance function of its derivative field is given by

$$\mathbb{E}\Big[\frac{dT(x)}{dx}\frac{dT(y)}{dy}\Big] = \frac{\partial^2 R(x,y)}{\partial x \partial y}.$$

We define the integration of a random field as the limit of Riemann sum. Let $\cup_{i=1}^n \mathcal{M}_i$ be a partition of $\mathcal{M}$, i.e. $\mathcal{M} = \cup_{i=1}^n \mathcal{M}_i$ and $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset$ if $i \neq j$. Let $x_i \in \mathcal{M}_i$ and $\mu(\mathcal{M}_i)$ be the volume of $\mathcal{M}_i$. Then we define the integration of field T as

$$\int_{\mathcal{M}} T(x)\, dx = \lim \sum_{i=1}^n T(x_i)\mu(\mathcal{M}_i),$$

where the limit is taken as $\mu(\mathcal{M}_j) \to 0$ for all j. When we integrate a Gaussian field, it is the limit of a linear combination of Gaussian random variables so it is again a Gaussian random variable. In general, any linear operation on Gaussian fields will result in Gaussian fields.

Other Fields. As in the case of Gaussian random variables, we can use Gaussian fields to construct new random fields such as χ^2, t, F and Hotelling's T^2 fields, all of which are extensively studied [25, 134, 133, 128]. For example, the χ^2-field with m degrees of freedom is defined as

$$T(x) = \sum_{i=1}^m X_i^2(x),$$

where $X_1, \cdots, X_m$ are independent, identically distributed Gaussian fields with zero mean and unit variance. Similarly, we can define t and F fields as well as Hotelling's T^2 field. The Hotelling's T^2-statistic has been widely used in detecting morphological changes in DBM [25, 36, 50, 69, 116]. In particular, [25] derived the excursion probability of the Hotelling's T^2-field and applied it to detect gender specific morphological differences.

Inference on Fields. Statistical inferences on random fields have been usually based on the series expansion of the form:

$$T(x) = \sum_{i=1}^\infty Z_i \phi_i(x), \tag{2.1}$$

where $\phi_i(x)$ are basis functions and Z_i are random variables. For a Gaussian random field, the most well known series expansion is called the *Karhunen-Loeve expansion* [42, 75, 135]. The inference on the expansion (2.1) is usually done on the realizations of coefficients Z_i [13]. In brain imaging, inference has been based on the extrema distributions of T [5, 76] which is given by

$$P\Big(\sup_{x\in\mathcal{M}} T(x) > h\Big).$$

This gives a localized inference in a sense that the thresholded region is identified as signal. On the other hand, the global inference can be done by integrating the field over the region of interest $\mathcal{M}$, i,e.

$$\int_{\mathcal{M}} T(x)\, d\mu(x),$$

which collapses the random field into a single random variable simplifying the inference. Details on the statistical inference will be given in the subsequent sections.

2.2 Multiple Comparisons

Given functional measurement Y, we have model

$$Y(x) = \mu(x) + \epsilon(x)$$

where μ is unknown signal to be estimated and ϵ is a zero mean unit variance Gaussian field. We further assume $x \in \mathcal{M} \subset \mathbb{R}^n$. In brain imaging, one of the most important problem is that of signal detection, which can be stated as the problem of identifying the regions of statistically significance. So it can be formulated as an inference problem

$$H_0 : \mu(x) = 0 \text{ for all } x \in \mathcal{M} \text{ vs. } H_1 : \mu(x) > 0 \text{ for some } x \in \mathcal{M}.$$

Let

$$H_0(x) : \mu(x) = 0$$

at a fixed point x. Then the null hypothesis H_0 is a collection of multiple hypotheses $H_0(x)$ over all x. Therefore, we have

$$H_0 = \bigcap_{x \in \mathcal{M}} H_0(x).$$

We may assume that $\mathcal{M}$ is the region of interest consisting of the finite number of voxels. We also have the corresponding point-wise alternate hypothesis

$$H_1(x) : \mu(x) > 0$$

and the alternate hypothesis H_1 is constructed as

$$H_1 = \bigcup_{x \in \mathcal{M}} H_0(x).$$

If we use Z-statistic as a test statistic, for instance, we will reject each $H_0(x)$ if $Z > h$ for some threshold h. So at each fixed x, for level $\alpha = 0.05$ test, we need

to have $h = 1.64$. However, if we threshold at $\alpha = 0.05$, 5% of observations are false positives. Note that the false positives are pixels where we are incorrectly rejecting $H_0(x)$ when it is actually true. However, this is the false positives related to testing $H_0(x)$. For determining the true false positives associated with testing H_0, we need to account for multiple comparisons. The type-I error is the probability of rejecting the null hypothesis (there is no signal) when the alternate hypothesis (there is signal) is true. The type-I error is also called the *family-wise error rate* (FWER) and given by

$$\begin{aligned} \alpha &= P(\text{ reject } H_0 \mid H_0 \text{ true }) \\ &= P(\text{ reject some } H_0(x) \mid H_0 \text{ true }) \\ &= P\Big(\bigcup_{x \in \mathcal{M}} \{Y(x) > h\} \;\Big|\; \mathbb{E}Y = 0 \Big). \end{aligned} \tag{2.2}$$

Unfortunately, $Y(x)$ is correlated over x and it makes the computation of type-I error almost intractable for random fields other than Gaussian.

Bonferroni Correction. One standard method for dealing with multiple comparisons is to use the Bonferroni correction. Note that the probability measure is additive so that for any event E_j, we have

$$P\Big(\bigcup_{j=1}^{\infty} E_j \Big) \leq \sum_{j=1}^{\infty} P(E_j).$$

This inequality is called Bonferroni inequalities and it has been used in the construction of simultaneous confidence intervals and multiple comparisons when the number of hypotheses are small. From (2.2), we have

$$\alpha = P\Big(\bigcup_{x \in \mathcal{M}} \{Y(x) > h\} \;\Big|\; \mathbb{E}Y = 0 \Big) \tag{2.3}$$

$$\leq \sum_{x \in \mathcal{M}} P\big(Y(x_j) > h \mid \mathbb{E}Y = 0\big) \tag{2.4}$$

So by controlling each type-I error separately at

$$P\big(Y(x_j) > h \mid \mathbb{E}Y = 0\big) < \frac{\alpha}{\#\mathcal{M}}$$

we can construct the correct level α test. Here $\#\mathcal{M}$ is the number of voxels.

The problem with the Bonferroni correction is that it is too conservative. The Bonferroni inequality (2.4) becomes exact when the measurements across voxels are all independent, which is unrealistic. Since the measurements are expected to be strongly correlated across voxels, we have highly correlated statistics. So in a sense, we have less number of comparisons to make.

2.3 Random Fields Theory

We can obtain less conservative estimate for (2.2) using the random field theory. Assuming $\mathbb{E}Y = 0$, we have

$$\begin{aligned}
\alpha(h) &= P\Big(\bigcup_{x\in\mathcal{M}} \{Y(x) > h\}\Big) \\
&= 1 - P\Big(\bigcap_{x\in\mathcal{M}} \{Y(x) \le h\}\Big) \\
&= 1 - P\Big(\sup_{x\in\mathcal{M}} Y(x) \le h\Big) \\
&= P\Big(\sup_{x\in\mathcal{M}} Y(x) > h\Big). \qquad (2.5)
\end{aligned}$$

In order to construct the α-level test corresponding to H_0, we need to know the distribution of the supremum of the field Y. The corresponding p-value based on the supremum of the field, i.e. $\sup_{x\in\mathcal{M}} Y$, is called the *corrected p-value* to distinguish it from the usual p-value obtained from the statistic Y. Note that the p-value is the smallest α-level at which the null hypothesis H_0 is rejected.

Analytically computing the exact distribution of the supremum of random fields is hard. If we denote $Z = \sup_{x\in\mathcal{M}} Y(x)$ and F_Z to be the cumulative distribution of Z, for the given given $\alpha = 0.05$, we can compuate $h = 1 - F_Z^{-1}(\alpha)$. Then the region of statistically significant signal is localized as $\{x \in \mathcal{M} : Y(x) > h\}$.

The distribution of supremum of Brownian motion is somewhat simple due to its independent increment properties. However, for smooth random field, it is not so straightforward. Read [3] for an overview of computing the distribution of the supremum of smooth fields.

Consider 1D smooth stationary Gaussian random process $Y(x), x \in \mathcal{M} = [0, 1] \subset \mathbb{R}$. Let N_h to be the number of times Y crosses over h from below (called upcrossing) in $[0, 1]$. Then we have

$$\begin{aligned}
P\Big(\sup_{x\in[0,1]} Y(x) > h\Big) &= P(N_h \ge 1 \text{ or } Y(0) > h) \\
&\le P(N_h \ge 1) + P(Y(0) > h) \\
&\le \mathbb{E}N_h + P(Y(0) > h).
\end{aligned}$$

If R is the covariance function of the field Y, we have

$$R(0) = \sigma^2 = \mathbb{E}Y^2(x).$$

It can be shown that from Rice formula [4, 101],

$$\mathbb{E}N_h = \frac{1}{\pi}\left(\frac{-R''(0)}{R(0)}\right)^{1/2} \exp\left(\frac{h^2}{2\sigma^2}\right).$$

Also note that $P(Y(0) > h) = 1-\Phi(\frac{h}{\sigma})$ where Φ is the cumulative distribution function of the standard normal. Then from the inequality that bounds the cumulative distribution of the standard normal [45], we have

$$\Big(1 - \frac{\sigma^2}{h^2}\Big)\frac{\sigma}{\sqrt{2\pi}h}e^{-h^2/2\sigma^2} \leq 1 - \Phi\Big(\frac{h}{\sigma}\Big) \leq \frac{\sigma}{\sqrt{2\pi}h}e^{-h^2/2\sigma^2}$$

So

$$P\Big(\sup_{x\in[0,1]} Y(x) > h\Big) \leq \Big[c_1 + \frac{c_2}{\sqrt{2\pi}h}\Big]e^{-h^2/2\sigma^2}$$

for some c_1 and c_2. In fact we can show that

$$P\Big(\sup_{x\in[0,1]} Y(x) > h\Big) = \Big[c_1 + \frac{c_2}{h} + O(h^{-2})\Big]e^{-h^2/2\sigma^2}.$$

Poisson Clumping Heuristic. To extend the Rice formula to higher dimension, we need a different mathematical machinery. For this method to work, the random field Y to be sufficiently smooth and isotropic. The smoothness of a random field corresponds to the random field being differentiable. There are very few cases for which exact formulas for the excursion probability (2.5) is known [2]. For this reason, approximating the excursion probability is necessary for most cases.

From the *Poisson clumping heuristic* [6],

$$P\Big(\sup_{x\in\mathcal{M}} Y(x) < h\Big) \approx \exp\bigg(-\frac{\|\mathcal{M}\|}{\mathbb{E}\|A_h\|}P\big(Y(x) \geq h\big)\bigg),$$

where $\|\cdot\|$ is the Lebesgue measure of a set and the random set

$$A_h = \{x \in \mathcal{M} : Y(x) > h\}$$

is called the *excursion set* above the threshold h. This approximation involves unknown $\mathbb{E}\|A_h\|$, which is the mean clump size of the excursion set. The distribution of $\|A_h\|$ has been estimated for the case of Gaussian [6], χ^2, t and F fields [23] but for general random fields, no approximation is available yet.

2.4 Euler Characteristic Approach

An alternate approximation based on the expected Euler characteristic (EC) of A_h is also available. The Euler characteristic approach reformulates the

geometric problem as a topological problem. Read [1], [26], [24], [113] and [129] for an overview of the Euler characteristic method.

For sufficiently high threshold h, it is known that

$$P\Big(\sup_{x\in\mathcal{M}} Y(x) > h\Big) \approx \mathbb{E}\chi(A_h) = \sum_{d=0}^{N} \mu_d(\mathcal{M})\rho_d(h) \tag{2.6}$$

where $\mu_d(\mathcal{M})$ is the d-th Minkowski functional or *intrinsic volume* of $\mathcal{M}$ and ρ_d is the d-th Euler characteristic (EC) density of Y [130]. For details on intrinsic volume, read [105]. The expansion (2.6) also holds for non-isotropic fields but we will not pursue it any further. Compared to other approximation methods such as the Poisson clump heuristic and the tube formulae, the advantage of using the Euler characteristic formulation is that a simple exact expression can be found for $\mathbb{E}\ \chi(A_h)$. Figure **??** and Figure **??** show how $\chi(A_h)$ and $\mathbb{E}\ \chi(A_h)$ change as h increases.

Intrinsic volume. The d-th intrinsic volume of $\mathcal{M}$ is a generalization of d-dimensional volume. Note that $\mu_0(\mathcal{M})$ is the Euler characteristic of $\mathcal{M}$. $\mu_N(\mathcal{M})$ is the volume of $\mathcal{M}$ while $\mu_{N-1}(\mathcal{M})$ is half the surface area of $\mathcal{M}$. There are various techniques for computing the intrinsic volume [113]. The methods depend on the smoothness of the underlying manifold $\mathcal{M}$. For a solid sphere with radius r, the intrinsic volumes are $\mu_0 = 1$, $\mu_1 = 4r$, $\mu_2 = 2\pi r^2$, $\mu_3 = \frac{4}{3}\pi r^3$. For a 3D box of size $a \times b \times c$, the intrinsic volumes are $\mu_0 = 1, \mu_1 = a + b + c, \mu_2 = ab + bc + ac, \mu_3 = abc$. In general, the intrinsic volume can be given in terms of a curvature matrix. Let $K_{\partial\mathcal{M}}$ be the curvature matrix of $\partial\mathcal{M}$ and $\mathrm{detr}_d(K_{\partial\mathcal{M}})$ be the sum of the determinant of all $d \times d$ principal minors of $K_{\partial\mathcal{M}}$. For $d = 0, \cdots, N-1$ the Minkowski functional $\mu_d(\mathcal{M})$ is defined as

$$\mu_d(\mathcal{M}) = \frac{\Gamma(\frac{N-i}{2})}{2\pi^{\frac{N-i}{2}}} \int_{\partial\mathcal{M}} \mathrm{detr}_{N-1-d}(K_{\partial\mathcal{M}})\, dA,$$

and $\mu_N(\mathcal{M}) = \|\mathcal{M}\|$, the Lebesgue measure of $\mathcal{M}$.

For nonregular jagged shapes such as the 2D corpus callosum shape $\mathcal{M}$, the intrinsic volume can be estimated in the following fashion. Treating pixels inside $\mathcal{M}$ as points on a lattice, let V be the number of vertices that forms the corners of pixels, E be the number of edges connecting each adjacent lattice points and F be the number of faces formed by four connected edges. We assume the distance between the adjacent lattice points is δ in all directions. Then $\mu_0 = V - E + F, \mu_1 = (E - 2F)\delta, \mu_2 = F\delta^2$ [132, 30]. To find the number of edges and pixels contained in $\mathcal{M}$, we start from an initial face (pixel) somewhere in the corpus callosum and add one face at a time while counting the additional edges and faces. In this fashion, we can grow a graph that will eventually contains all the pixels that form the corpus callosum. Figure **??** shows few possible configurations of adding a pixel (black) to the existing graph (gray pixels). A similar approach for computing the intrinsic

volume for jagged irregular shapes has been implemented in FMRISTAT package (http://www.math.mcgill.ca/keith/fmristat).

EC-density. The d-th EC-density is given by

$$\rho_d(h) = \mathbb{E}\big[(Y > h)\det(-\ddot{Y}_d)|\dot{Y}_d = 0\big]P(\dot{Y}_d = 0),$$

where dot notation indicates partial differentiation with respect to the first d components. The subscript d represents the first d components of Y. Computation of this conditional expectation is nontrivial other than Gaussian fields. For zero mean and unit variance Gaussian field Y, we have for instance

$$\begin{aligned}
\rho_0 &= P(Y > h) = 1 - \Phi(h) \\
\rho_1 &= \lambda^{1/2}\frac{e^{-h^2/2}}{\sqrt{2\pi}} \\
\rho_2 &= \lambda h\frac{e^{-h^2/2}}{\sqrt{2\pi}} \\
\rho_3 &= \lambda^{3/2}(h^2-1)\frac{e^{-h^2/2}}{\sqrt{2\pi}},
\end{aligned}$$

where λ measures the smoothness of fields, defined as the variance of the derivative of component of Y. The exact expression for the EC density ρ_d is available for other random fields such as t, χ^2, F fields [128], Hotelling's T^2 fields [25] and scale-space random fields [108]. In each case, the EC density ρ_d is proportional to $c^{\frac{d}{2}}$ and it changes depending on the smoothness of the field.

If $X_1, \cdots, X_\alpha, Y_1, \cdots, Y_\beta$ are i.i.d. stationary zero mean unit variance Gaussian fields. Then F-field with α and β degrees of freedom is given by

$$F(x) = \frac{\sum_{j=1}^{\alpha} X_j^2(x)/\alpha}{\sum_{j=1}^{\beta} Y_j^2(x)/\beta}.$$

To avoid singularity, we need to assume the total degrees of freedom $\alpha+\beta \gg N$ to be sufficiently larger than the dimension of space [128]. The EC-density for F-field is then given by

$$\begin{aligned}
\rho_0 &= \int_h^\infty \frac{\Gamma(\frac{\alpha+\beta}{2})}{\Gamma(\frac{\alpha}{2})\Gamma(\frac{\beta}{2})}\frac{\alpha}{\beta}\left(\frac{\alpha x}{\beta}\right)^{\frac{(\alpha-2)}{2}}\left(1+\frac{\alpha x}{\beta}\right)^{-\frac{(\alpha+\beta)}{2}} dx, \\
\rho_1 &= \lambda^{1/2}\frac{\Gamma(\frac{\alpha+\beta-1}{2})2^{\frac{1}{2}}}{\Gamma(\frac{\alpha}{2})\Gamma(\frac{\beta}{2})}\left(\frac{\alpha h}{\beta}\right)^{\frac{(\alpha-1)}{2}}\left(1+\frac{\alpha h}{\beta}\right)^{-\frac{(\alpha+\beta-2)}{2}}, \\
\rho_2 &= \lambda\frac{\Gamma(\frac{\alpha+\beta-2}{2})}{\Gamma(\frac{\alpha}{2})\Gamma(\frac{\beta}{2})}\left(\frac{\alpha h}{\beta}\right)^{\frac{(\alpha-2)}{2}}\left(1+\frac{\alpha h}{\beta}\right)^{-\frac{(\alpha+\beta-2)}{2}} \\
&\times \left[(\beta-1)\frac{\alpha h}{\beta} - (\alpha-1)\right].
\end{aligned}$$

If the random field Y is given as the convolution of a smooth kernel $K_h(x) = K(x/h)/h^N$ with a white Gaussian noise [108, 131], the covariance matrix of $\dot{Y} = dY/dx$ is given by

$$\mathbf{Var}(\dot{Y}) = \frac{\int_{\mathbb{R}^N} \dot{K}(\frac{x}{h})\dot{K}^t(\frac{x}{h})\, dx}{h^2 \int_{\mathbb{R}^N} K^2(\frac{x}{h})\, dx}.$$

Applying it to a Gaussian kernel $K(x) = (2\pi)^{-n/2} e^{-\|x\|^2/2}$ gives $c = \mathbf{Var}(\dot{Y}_1) = 1/(2h^2)$. In terms of the full width at half maximum (FWHM) of the kernel K_h, $c = 4 \ln 2/\text{FWHM}^2$.

2.5 Power and Sample Size

Given the null hypothesis H_0 and the alternate hypothesis H_1 on parameters of an underlying statistical model, let us define what is the power of a test statistic. The probabilities of type-I (α) and type-II (β) errors are defined respectively as:

$$\begin{aligned} \alpha &= P(\text{Type I error}) \\ &= P(\text{reject } H_0 \mid H_0 \text{ true}). \end{aligned}$$

$$\begin{aligned} \beta &= P(\text{Type II error}) \\ &= P(\text{not reject } H_0 \mid H_0 \text{ false }) \\ &= 1 - P(\text{reject } H_0 \mid H_1 \text{ true}). \end{aligned}$$

Then the *power of* the test is defined to be $1 - \beta$.

$$\text{Power } = P(\text{reject } H_0 \mid H_1 \text{ ture}).$$

When the test procedure has the power of 0.8, it implies that we can correctly reject the null hypothesis H_0 80% of the time when the alternate hypothesis H_1 is true. The sample size computation is then based on power. We are usually interested in how many samples we need to obtain a specific power level.

Consider two samples $X_1, \cdots, X_{n_1} \sim N(\mu_1, \sigma^2)$ and $Y_1, \cdots, Y_{n_2} \sim N(\mu_2, \sigma^2)$. We are interested in testing

$$H_0 : \mu_1 - \mu_2 = 0 \text{ vs. } H_1 : \mu_1 - \mu_2 = c\sigma \neq 0.$$

The constant c represent the mean difference with respect to the standard deviation.

For a test statistic, we use the t-statistic with the equal variance assumption:

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{1/n_1 + 1/n_2}}, \tag{2.7}$$

where $\bar{X}$ and $\bar{Y}$ are the sample means and S_p^2 is the pooled sample variance. If the sample variance of the i-th group is denoted by S_i^2, the pooled sample variance is given by

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

It can be shown that the test statistic T is distributed as

$$T \sim t_{n_1+n_2-2},$$

the student t-distribution with $n_1 + n_2 - 2$ degrees of freedom.

For computing the power, the α-level has to be specified first. Under H_0, the rejection region corresponding to the α-level is given by

$$\frac{|\bar{X} - \bar{Y}|}{S_p\sqrt{1/n_1 + 1/n_2}} > t_{n_1+n_2-2,\alpha/2},$$

where $t_{n_1+n_2-2,\alpha/2}$ is the quantile satisfying

$$P(T \geq t_{n_1+n_2-2,\alpha/2}) = \alpha/2.$$

Here we assumed $\sigma \approx S_p$.

Under H_1, $X_i \sim N(\mu_1, \sigma^2)$ and $Y_i \sim N(\mu_1 + c\sigma, \sigma^2)$. So it follows

$$T = \frac{\bar{X} - \bar{Y} - c\sigma}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t_{n_1+n_2-2}.$$

Then the power is given by

$$\begin{aligned} \text{Power} &= P\Big(-t_{n_1+n_2-2,\alpha/2} < \frac{\bar{X} - \bar{Y}}{S_p\sqrt{1/n_1 + 1/n_2}} < t_{n_1+n_2-2,\alpha/2} \Big| H_1 \Big) \\ &= P\Big(-t_{n_1+n_2-2,\alpha/2} - \frac{c}{\sqrt{1/n_1 + 1/n_2}} < T < t_{n_1+n_2-2,\alpha/2} - \frac{c}{\sqrt{1/n_1 + 1/n_2}} \Big). \end{aligned}$$

For sufficiently large n_1 and n_2, we may assume $T \sim N(0, 1)$. Let Φ be the cumulative distribution for the standard normal distribution. Then our power is approximated as

$$\text{Power}(n_1, n_2) = 1 + \Phi\Big(-z_{\alpha/2} - \frac{c}{\sqrt{1/n_1 + 1/n_2}} \Big) - \Phi\Big(z_{\alpha/2} - \frac{c}{\sqrt{1/n_1 + 1/n_2}} \Big).$$

Assuming $n = n_1 = n_2$, we can plot the power as a function of the sample

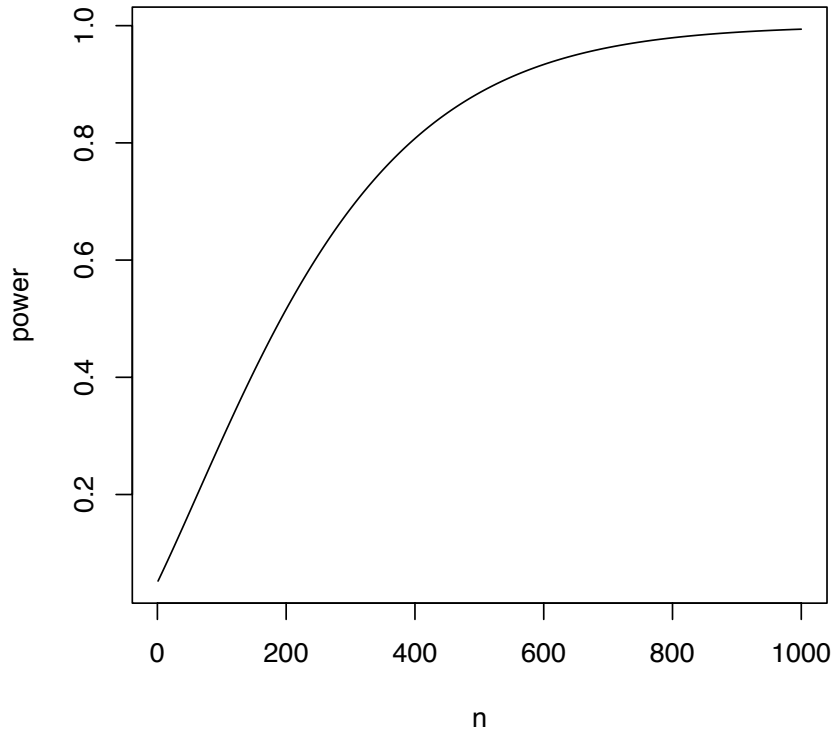

FIGURE 2.1
Power vs. sample size for the two sample t test.

size. For example, in order to obtain power of 0.8 for a $\alpha = 0.05$ test in differentiating the difference $\mu_1 - \mu_2 = 0.2\sigma$, we need $n = n_1 = n_2 = 393$. See Figure 2.1 for the power vs. sample size plot.

So far, the power computation is based on samples for scalar measurements. Assume we have two functional measurements $X_1(t), \cdots, X_{n_1}(t)$ and $Y_1(t), \cdots, Y_{n_2}(t)$ over continuous index $t \in \mathcal{M}$. X_i and X_j can be then modeled as random fields over $\mathcal{M}$. At each fixed t, we have the same test statistic $T(t)$ given in (2.7).

The usual point-wise hypotheses are given by

$$H_0(t) : \mu_1(t) - \mu_2(t) = 0 \text{ vs. } H_1(t) : \mu_1(t) - \mu_2(t) = c\sigma > 0$$

for each fixed t. Instead of the point-wise inference, what we need is a global

inference for the whole parameter space $\mathcal{M}$. The usual global hypotheses accounting for multiple comparisons are then given by

$$J_0 : \mu_1(t) - \mu_2(t) = 0 \text{ for all } t \in \mathcal{M} \text{ vs. } J_1 : \mu_1(t) - \mu_2(t) = c\sigma > 0 \text{ for some } t \in \mathcal{M}.$$

The relationship between the point-wise hypotheses $H_0(t), H_1(t)$ and the global hypotheses J_0, J_1 are

$$J_0 = \bigcap_{t \in \mathcal{M}} H_0(t), \; J_1 = \bigcup_{t \in \mathcal{M}} H_1(t).$$

Type-I error. In order to compute the power over $\mathcal{M}$, it is necessary to determine the type-I error first. Note that we reject J_0 if $T(t) > h$ for some thresholding h for all $t \in \mathcal{M}$. This is equivalent to the event $\sup_{t \in \mathcal{M}} T(t) > h$. Hence, the type-I error computation requires knowing the distribution of the random variable $\sup_{t \in \mathcal{M}} T(t)$ which can be very involving. The type-I error over $\mathcal{M}$ is given by

$$\alpha = P\Big(\sup_{t \in \mathcal{M}} T(t) > t^*_\alpha \Big),$$

where t^*_α is the quantile corresponding to the random variable $\sup_{t \in \mathcal{M}} T(t)$. If necessary, the qauntile can be determined numerically using the permutation test.

Type-II error. Let

$$T(t) = \frac{\bar{X}(t) - \bar{Y}(t)}{S_p(t)\sqrt{1/n_1 + 1/n_2}}.$$

Under J_0, this is a t-random field with $n_1 + n + 2 - 2$ degrees of freedom. The rejection region of J_0 corresponding to the α level is given by

$$\sup_{t \in \mathcal{M}} T(t) > t^*_\alpha.$$

Under J_1, we have

$$X_i(t) \sim N(\mu_1, \sigma^2) \text{ and } Y_i(t) \sim N(\mu_1 + c\sigma, \sigma^2)$$

for some region $\mathcal{M}_0 \subset \mathcal{M}$. In other region $\mathcal{M}/\mathcal{M}_0$, we simply have

$$X_i(t), Y_i(t) \sim N(\mu_1, \sigma^2).$$

To compute the power over the whole space $\mathcal{M}$, we need the conditional probability $P($ reject $J_0 \mid J_1$ true $)$. Conditioning on J_1 introduces two distinct possibilities ($\mathcal{M}_0$ and $\mathcal{M}/\mathcal{M}_0$) so the power computation over the whole $\mathcal{M}$ is slightly complicated. Figure 2.2 shows the event when J_1 is true. In $\mathcal{M}_0$, we have

$$T_0(t) = T(t) - \frac{c}{\sqrt{1/n_1 + 1/n_2}} \sim t_{n_1+n_2-2}$$

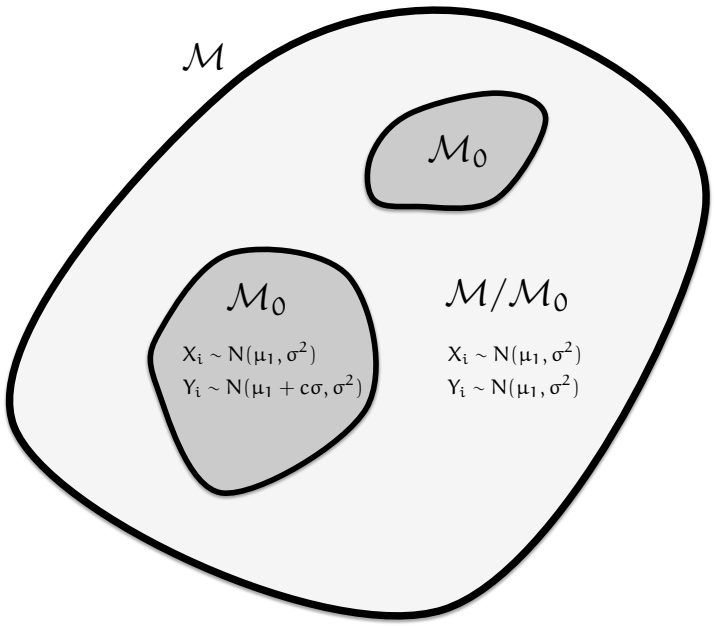

FIGURE 2.2
Under J_1, there exists a nonempty region $\mathcal{M}_0$ where the means of X_i and Y_i are different.

pointwisely. Here we assumed the equal variance field, i.e. $S_p(t) = \sigma$. On the other hand, in the region $\mathcal{M}/\mathcal{M}_0$, we have

$$T(t) \sim t_{n_1+n_2-2}$$

pointwisely.

Hence the overall power over $\mathcal{M}$ is given by

$$\mathcal{P} = P\Big(\max \Big[\sup_{t\in\mathcal{M}_0} T_0(t) + \frac{c}{\sqrt{1/n_1 + 1/n_2}}, \sup_{t\in\mathcal{M}/\mathcal{M}_0} T(t) \Big] > t^*_\alpha \Big).$$

Since the analytic derivation of the exactly probability is intractable, we will approximate the power by assuming the random fields T_0 restricted to $\mathcal{M}_0$ and T restricted to $\mathcal{M}/\mathcal{M}_0$ to be independent. The algebraic manipulation can show that the power is then approximately

$$\mathcal{P} = P\Big(\sup_{t\in\mathcal{M}_0} T_0(t) > t^*_\alpha - \frac{c}{\sqrt{1/n_1 + 1/n_2}} \Big) + P\Big(\sup_{t\in\mathcal{M}} T(t) > t^*_\alpha - \frac{c}{\sqrt{1/n_1 + 1/n_2}} \Big).$$

by bounding it. The algebraic manipulation can show that the power is

bounded between

$$P\Big(\sup_{t\in\mathcal{M}_0} T_0(t) > t^*_\alpha - \frac{c}{\sqrt{1/n_1+1/n_2}}\Big) < \mathcal{P} \leq P\Big(\sup_{t\in\mathcal{M}} T(t) > t^*_\alpha - \frac{c}{\sqrt{1/n_1+1/n_2}}\Big).$$

For sufficiently large n_1 and n_2, we can approximate the t-fields T_0 and T using the Gaussian field Z with zero mean and unit variance. The tail distribution of the supremum of Z field in $\mathbb{R}^2$ is given approximately as [133]:

$$\Psi(t,\mathcal{M}) = P\Big(\sup_{t\in\mathcal{M}} Z(t) > t\Big) = \frac{\mu(\mathcal{M})}{FWHM^2}\frac{4\ln 2}{(2\pi)^{3/2}} t\exp(-t^2/2),$$

where $\mu(\mathcal{M})$ is the surface area of and FWHM is the full-width-at-half-maximum of signal or smoothing kernel. Hence the power is bounded by

$$\Psi\Big(t^*_\alpha - \frac{c}{\sqrt{1/n_1+1/n_2}}, \mathcal{M}_0\Big) \leq \mathcal{P} \leq \Psi\Big(t^*_\alpha - \frac{c}{\sqrt{1/n_1+1/n_2}}, \mathcal{M}\Big).$$

For instance, in the cortical thickness analysis, often used FWHM is $20mm$ kernel and the area of the outer cortical surface is $302180mm^2$ [34]. Assuming $n = n_1 = n_2$, we can plot the power as a function of the sample size (Figure 2.3). In order to obtain the power of 0.8 for a $\alpha = 0.05(corrected)$ test in differentiating $\mu_1 - \mu_2 = 0.2\sigma$, we need $n_1 = n_2$ in each group.

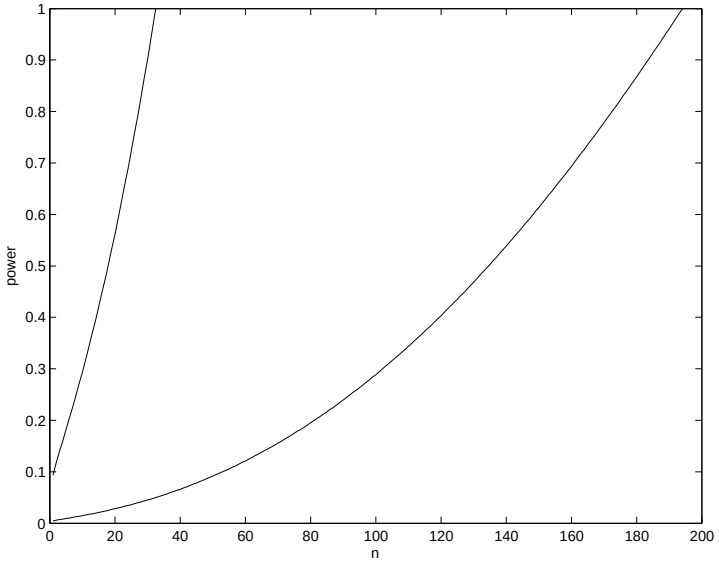

FIGURE 2.3
Power function bounded by the lower and upper bounds.

3

Spherical Harmonic Representation

NOTE: MATLAB COMPONENTS NOT ADDED YET.

There are extensive literature on local cortical shape modeling and analysis [32, 46, 70, 114, 118, 77, 80, 86]). The medial representation [95] has been also successfully used in modeling various subcortical structures including the cross sectional images of the corpus callosum [68], hippocampus and amygdala complex [110], ventricle and brain stem [95]. In the medial representation, the binary object is represented using the finite number of atoms and links that connect the atoms together to form a skeletal representation of the object. The medial representation is mainly used with the principal component analysis type of approach for shape classification and group comparison.

Unlike the medial representation, which is in a discrete representation, there is a continuous parametric approach called the spherical harmonic representation [53, 60, 72, 107]. The spherical harmonic representation has been mainly used as a data reduction technique for compressing global shape features into small number of coefficients. The main global geometric features are encoded in low degree coefficients while the noise will be in high degree spherical harmonics [60]. The method has been used to model various subcortical structures such as ventricles [53], hippocampi [107] and cortical surfaces [31]. The spherical harmonics have global support. So the spherical harmonic coefficients contain only the global shape features and it is not possible to directly obtain local shape information from the coefficients only. However, it is still possible to obtain local shape information by evaluating the representation at each fixed point, which gives the smoothed version of the coordinates of surfaces. In this fashion, the spherical harmonic representation can be viewed as mesh smoothing [31]. Instead of using the global basis of spherical harmonics, there have been attempts of using the local wavelet basis for parameterizing cortical surfaces [88, 136].

Other shape modeling approaches include distance transforms [78], deformation fields [85] obtained by warping individual substructures to a template, and the particle-based method [27]. A distance transform is a function that for each point in the image is equal to the distance from that point to the boundary of the object [56]. The distance map approach has been applied in classifying a collection of hippocampus [56]. The deformation fields based approach has been somewhat popular and has been applied to modeling whole 3D brain volume [9, 33, 50], cortical surfaces [34, 117], hippocampus [70] and

cingulate gyrus [38]. The particle-based method uses a nonparametric, dynamic particle system to simultaneously sample object surfaces and optimize correspondence point positions [27].

In this chapter, we presented the unified mathematical theory of the Fourier representation which encompasses the spherical harmonic representation as a special case in $\mathbb{R}^3$. The representation can be used for cortical surface parameterization, smoothing and registration in a unified Hilbert space framework. The weighted version of the Fourier representation is also developed to address many shortcomings of the traditional spherical harmonic representation [31]. The weighted version differs from the traditional spherical harmonic representation in many ways. Although the truncation of the series expansion in the spherical harmonic representation can be viewed as a form of smoothing, there is no direct equivalence to the full width at half maximum (FWHM) usually associated with kernel smoothing. So it is difficult to relate the unit of FWHM widely used in brain imaging to the degree of spherical harmonic representation. On the other hand, the weighted representation can easily relate to FWHM of smoothing kernel so we have a clear sense of how much smoothing we are performing beforehand. The traditional representation suffers from the Gibbs phenomenon (ringing artifacts) [51] that usually happens in representing rapidly changing or discontinuous data with smooth periodic basis. The weighted representation can substantially reduce the amount of Gibbs phenomenon by weighting the coefficients of the spherical harmonic expansion.

3.1 Fourier Representation

Consider a compact differentiable manifold $\mathcal{M} \in \mathbb{R}^d$ that will be our anatomical object of interest. Let $L^2(\mathcal{M})$ be the space of square integrable functions in $\mathcal{M}$ with inner product

$$\langle g_1, g_2 \rangle = \int_{\mathcal{M}} g_1(p) g_2(p)\, d\mu(p), \tag{3.1}$$

where μ is the Lebegue measure such that $\mu(\mathcal{M})$ is the total volume of $\mathcal{M}$. The norm $\|\cdot\|$ is defined as

$$\|g\| = \langle g, g \rangle^{1/2}.$$

The partial differential operator $\mathcal{L}$ is *self-adjoint* if

$$\langle g_1, \mathcal{L} g_2 \rangle = \langle \mathcal{L} g_1, g_2 \rangle$$

for all $g_1, g_2 \in L^2(\mathcal{M})$. The eigenvalues λ_j and eigenfunctions ψ_j of the operator $\mathcal{L}$ are obtained by solving

$$\mathcal{L}\psi_j = \lambda_j \psi_j. \tag{3.2}$$

Without the loss of generality, we can order eigenvalues

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$$

and make the eigenfunctions to be orthonormal with respect to the inner product (3.1).

Let $\mathcal{H}_k$ be the subspace

$$\mathcal{H}_k = \{\sum_{j=0}^{k} \beta_j \psi_j(p) : \beta_j \in \mathbb{R}\} \subset L^2(\mathcal{M}),$$

which is spanned by the finite number of basis up to degree k. We are interested in finding a function $h \in \mathcal{H}_k$ that is the closest to f in L_2-norm. Obviously, from the property of Hilbert space $L^2(\mathcal{M})$, we have

$$\sum_{j=0}^{k} f_j \psi(p) = \arg\min_{h \in \mathcal{H}} \|f - h\|^2,$$

where $f_j = \langle f, \psi_j \rangle$ are Fourier coefficients. There are three main methods for computing Fourier coefficients. The first method numerically integrates the Fourier coefficients over a high resolution triangle mesh [28]. Although this approach is the simplest to implement numerically and possibly the most accurate, the computation is extremely slow, due to the brute force nature of the technique. This is not a recommended approach. The second method is based on the fast Fourier transform (FFT) [20, 60]. The drawback of FFT is the need for a predefined regular grid system so if the mesh topology is different for different surfaces, a time consuming interpolation is needed. Cortical meshes obtained from FreeSurfer [46] produces topologically different meshes for different subjects so FFT is also not recommended. The third method is based on solving a system of linear equations [53, 107, 106] in a least squares fashion. This is the most widely used numerical technique in the spherical harmonic representation literature. However, the direct application of the least squares estimation is not desirable when the size of the linear equation is extremely large. For extremely large least squares problems, iterative strategies such as the *iterative residual fitting* (IRF) is required [31].

Suppose f is observed at the finite number of points $p_1, \cdots, p_n \in \mathcal{M}$. Then we wish to find $h \in \mathcal{H}_k$ that minimizes the sum of the squared distance

$$\|f - h\|^2 \approx \sum_{i=1}^{n} \Big[f(p_i) - \sum_{j=0}^{k} \beta_j \psi_j(p_i) \Big]^2. \tag{3.3}$$

The minimum of (3.3) is obtained when

$$f(p_i) = \sum_{j=0}^{k} \beta_j \psi_j(p_i), i = 1, \cdots, n \tag{3.4}$$

The equation (3.4) is referred as the *normal equation* and is usually solved by matrix inversion as follows. Let $f = (f(p_1), \cdots, f(p_n))'$ and $\beta = (\beta_0, \cdots, \beta_k)'$. Also let

$$\Psi = \begin{bmatrix} \psi_0(p_1) & \cdots & \psi_k(p_1) \\ \vdots & \ddots & \vdots \\ \psi_0(p_n) & \cdots & \psi_k(p_n) \end{bmatrix}$$

be a $n \times (k+1)$ matrix consisting of basis functions evaluated at mesh vertices. Then (3.4) can be rewritten in the following matrix form:

$$f = \Psi\beta. \tag{3.5}$$

The solution of the matrix equation is

$$\beta = (\Psi'\Psi)^{-}\Psi' f, \tag{3.6}$$

where $(\Psi'\Psi)^{-}$ is the generalized inverse. The problem with this widely used formulation is that the size of the matrix Ψ can be fairly large and for very large n and k. So it may become impractical to perform matrix operation (3.6) directly. This is mainly true for FreeSurfer [46] which produces more than $200,000$ nodes for each cortical hemisphere. This computational bottleneck can be overcome by breaking the least squares problem in the subspace $\mathcal{H}_k$ into smaller subspaces using the IRF-algorithm [106, 31].

3.2 Weighted Fourier Representation

The *weighed Fourier representation* generalizes the usual Fourier representation with additional exponential weights. This new representation is both a global hierarchical parameterization and an explicit data smoothing technique formulated as a solution to a self-adjoint partial differential equation (PDE). The exponentially decaying weights make the representation converges faster and reduce the Gibbs phenomenon (ringing artifacts) significantly [51]. When the self-adjoint operator $\mathcal{L}$ is the Laplace-Beltrami operator, the representation becomes heat kernel smoothing [31].

Consider a Cauchy problem

$$\frac{\partial g}{\partial \sigma} + \mathcal{L}g = 0, g(p, \sigma = 0) = f(p). \tag{3.7}$$

The initial functional data $f(p)$ can be further stochastically modeled as

$$f(p) = \eta(p) + \epsilon(p), \tag{3.8}$$

where ϵ is a mean zero Gaussian random field and η is the unknown signal

to be estimated. The partial differential equation (3.7) diffuses initial data f over time and the solution is given as the estimate for η. The time σ controls the amount of smoothing and will be termed as the *bandwidth*. Using the eigenfunctions (3.2), the unique solution to equation (3.7) is given by

$$g(p,\sigma) = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} f_j \psi_j(p) \tag{3.9}$$

with Fourier coefficients $f_j = \langle f, \psi_j \rangle$. For each fixed σ, g has expansion

$$g(p,\sigma) = \sum_{j=0}^{\infty} c_j(\sigma)\psi_j(p). \tag{3.10}$$

Substitute equation (3.10) into (3.7). Then we obtain

$$\frac{\partial c_j(\sigma)}{\partial \sigma} + \lambda_j c_j(\sigma) = 0. \tag{3.11}$$

The solution of equation (3.11) is given by $c_j(\sigma) = b_j e^{-\lambda_j \sigma}$. So we have solution

$$g(p,\sigma) = \sum_{j=0}^{\infty} b_j e^{-\lambda_j \sigma} \psi_j(p).$$

At $\sigma = 0$, we have

$$g(p,0) = \sum_{j=0}^{\infty} b_j \psi_j(p) = f(p).$$

The coefficients b_j must be the Fourier coefficients $\langle f, \psi_j \rangle$ and this proves our claim. The solution (3.9) decreases exponentially as time σ increases and smoothes out high spatial frequency noise much faster than low frequency noise. This is the basis of many of PDE-based image smoothing methods. Partial differential equations involving self-adjoint linear partial differential operators such as the Laplace-Beltrami operator or iterated Laplacian have been widely used in medical image analysis as a way to smooth either scalar or vector data along anatomical boundaries [7, 20, 21, 35]. These methods directly solve the PDE using standard numerical techniques such as the finite difference method or the finite element method. However, the main problem with directly solving PDE is the numerical instability and the complexity of setting up the numerical scheme.

Heat kernel K_σ is approximated linearly using Gaussian kernel in the tangent space. This process bounds to compound the linearization error. The linearization problem can be avoided if we can determine heat kernel precisely. Motivated by the solution (3.9), we define the *weighted Fourier representation* of f as

$$\sum_{j=0}^{\infty} e^{-\lambda_j \sigma} f_j \psi_j(p). \tag{3.12}$$

By rearranging the inner product in (3.12), we have

$$\begin{aligned}\sum_{j=0}^{\infty} e^{-\lambda_j \sigma} f_j \psi_j(p) &= \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} \psi_j(p) \int_{\mathcal{M}} f(q)\psi_j(q)\, d\mu(q) \\ &= \int_{\mathcal{M}} K_\sigma(p,q) f(q)\, d\mu(q)\end{aligned}$$

with the positive definite symmetric kernel K_σ given by

$$K_\sigma(p,q) = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} \psi_j(p)\psi_j(q).$$

This shows that the solution of the Cauchy problem (3.7) can be interpreted as kernel smoothing

$$K_\sigma * f = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} f_j \psi_j(p).$$

When the differential operator $\mathcal{L} = \Delta$, the Laplace-Beltrami operator, the Cauchy problem (3.7) becomes an isotropic diffusion equation. For this particular case, K_σ is called the *heat kernel* with bandwidth σ [32, 28]. For an arbitrary cortical manifold, the basis functions ψ_j can be computed and the exact shape of heat kernel can be determined numerically. Although it can be done by setting up a huge finite element method [98], this is not a trivial numerical computation. A simpler approach is to use the first order approximation of the heat kernel for small bandwidth and iteratively apply it up to the desired bandwidth [32].

The weighted Fourier representation can be reformulated as a kernel regression problem [44]. We restrict the function space $L^2(\mathcal{M})$ to a finite subspace that is more useful in numerical implementation. Let

$$\mathcal{H}_k = \{\sum_{j=0}^{k} \beta_j \psi_j(p) : \beta_j \in \mathbb{R}\}$$

be the subspace spanned by basis $\psi_0, \cdots, \psi_l$. We claim that the k-th degree expansion of (3.12) satisfies

$$\sum_{j=0}^{k} e^{-\lambda_j \sigma} f_j \psi_j = \arg\min_{h \in \mathcal{H}_k} \int_{\mathcal{M}} \int_{\mathcal{M}} K_\sigma(p,q) \Big| f(q) - h(p) \Big|^2 \, d\mu(p)\, d\mu(q).$$

This can be seen by letting $h = \sum_{j=0}^{k} \beta_j \psi_j(p)$. Let the inner integral be

$$I = \int_{\mathcal{M}} K_\sigma(p,q) \Big| f(q) - \sum_{j=0}^{k} \beta_j \psi(p) \Big|^2 \, d\mu(q).$$

Simplifying the expression, we obtain

$$I = \sum_{j=0}^{k} \sum_{j'=0}^{k} \psi_j(p)\psi_{j'}(p)\beta_j\beta_{j'} - 2K_\sigma * f(p) \sum_{j=0}^{k} \psi_j(p)\beta_j + K_\sigma * f^2.$$

Since I is an unconstrained positive semidefinite qudratic program (QP) in β_j, there is no unique global minimizer of I without additional linear constraints. Integrating I further with respect to $\mu(p)$, we collapses the QP to a positive definite QP, which yields a unique global minimizer as

$$\int_{\mathcal{M}} I \, d\mu(p) = \sum_{j=0}^{k} \beta_j^2 - 2 \sum_{j=0}^{k} e^{-\lambda_j \sigma} f_j \beta_j + \text{ const.}$$

The minimum of the above integral is obtained when all the partial derivatives with respect to β_j vanish, i.e.

$$\int_{\mathcal{M}} \frac{\partial I}{\partial \beta_j} \, d\mu(p) = 2\beta_j - 2e^{-\lambda_j \sigma} f_j = 0$$

for all j. Hence $\sum_{j=0}^{k} e^{-\lambda_j \sigma} f_j \psi_j$ is the unique minimizer in $\mathcal{H}_k$.

We can also show that the weighted spherical harmonic representation is related to previously available surface-based isotropic diffusion smoothing [7, 21, 34, 32]. When $\mathcal{L} = \Delta$, the weighted Fourier representation (3.9) is the solution of the isotropic heat diffusion. Then from the property of the generalized Fourier series, the finite expansion is the closest to the infinite series in $\mathcal{H}_k$ in the least squares fashion [104]. This can be formally stated as

$$\sum_{j=0}^{k} e^{-\lambda_j \sigma} f_j \psi_j = \arg \min_{h \in \mathcal{H}_k} \|h - h_0\|,$$

where h_0 is the solution to the isotropic heat diffusion

$$\frac{\partial h_0}{\partial \sigma} = \Delta h_0, \tag{3.13}$$

with the initial value condition $h_0(p, \sigma = 0) = f(p)$ in the manifold $\mathcal{M}$.

3.3 Iterative Residual Fitting Algorithm

We present an iterative technique for solving (3.5) for extremely large number of basis k. Decompose the subspace $\mathcal{H}_k$ into smaller subspaces as the direct sum:

$$\mathcal{H}_k = \mathcal{I}_0 \oplus \mathcal{I}_1 \cdots \oplus \mathcal{I}_k,$$

where subspace $\mathcal{I}_l$ is the the projection of $\mathcal{H}_k$ along the k-th basis. Other way of decomposing $\mathcal{H}_k$ is to use more than one basis for $\mathcal{I}_l$. For instance, for the collection of spherical harmonics Y_{lm}, at each degree l, there are $2l+1$ basis $Y_{l,-l}, \cdots, Y_{l,l}$. So we define $\mathcal{I}_l$ as the $2l+1$ dimensional subspace generated by all l-th degree spherical harmonics. Then the algorithm estimates the Fourier coefficients β_j in each subspace $\mathcal{I}_l$ iteratively from increasing the degree from 0 to k. Suppose we estimated the coefficients up to degree $l-1$ somehow. The estimated coefficients are denoted as $\widehat{\beta}_0, \cdots, \widehat{\beta}_{l-1}$. Then the residual r_{l-1} of the fit is given by

$$r_{l-1} = f - \sum_{j=0}^{l-1} \widehat{\beta}_j \psi_j. \tag{3.14}$$

At the next degree l, we estimate the coefficients β_l by minimizing the difference between the residual r_{l-1} and $\beta_l \psi_l$, i.e.

$$\widehat{\beta}_l = \arg\min_{\beta_l} \|r_{l-1} - \beta_l \psi_l\|^2.$$

The minimization is achieved in the least squares fashion with a smaller normal equation. Let $\Psi_l = (\psi_l(p_1), \cdots, \psi_l(p_n))'$. Then

$$\widehat{\beta}_l = (\Psi_l' \Psi_l)^{-1} \Psi_l' r_{l-1}.$$

The model for given functional data f is then $\sum_{j=0}^{l} \widehat{\beta}_j \psi_j$. In this fashion, the algorithm hierarchically builds the Fourier expansion from lower to higher degree. To speed up the computation, we can decompose $\mathcal{H}_k$ such that each subspace $\mathcal{I}_k$ is spanned by more than one basis if necessary. The iterative procedure presented here is refereed to as the *iterative residual fitting (IRF)* algorithm since we are iteratively fitting a linear equation to the residuals obtained from the previous iteration [31].

For spherical harmonics, due to multiplicity, there are $2l+1$ orthonormal basis corresponding to the l-th eigenvalue. So we can simultaneously estimate more than one coefficient at a time. The procedure is similar to estimating one coefficient at a time. Here we spell out the procedure for estimating the coefficients in the weighted Fourier representation. Suppose we have the normal equations

$$f(p_j) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)\sigma} f_{lm} Y_{lm}(p_j), \; j = 1, \cdots, n, \tag{3.15}$$

where $f_l m = \langle f, Y_{lm} \rangle$. We rewrite (3.15) in the matrix form as

$$\mathbf{F} = \underbrace{[\mathbf{Y}_0, e^{-1(1+1)\sigma} \mathbf{Y}_1, \cdots, e^{-k(k+1)\sigma} \mathbf{Y}_k]}_{\mathbf{Y}} \beta, \tag{3.16}$$

where the column vectors are $\mathbf{F} = [f(p_1), \cdots, f(p_n)]'$ and $\beta' =$

$(\beta'_0, \beta'_1, \cdots, \beta'_k)$ with $\beta'_l = (f_{l,-l}, f_{l,l})$. The length of the vector β is $1 + (2 \cdot 1 + 1) + \cdots + (2 \cdot k + 1) = (k+1)^2$. Each submatrix $\mathbf{Y}_l$ is given by

$$\mathbf{Y}_l = \begin{bmatrix} Y_{l,-l}(p_1), & \cdots & , Y_{l,l}(p_1) \\ \vdots & \ddots & \vdots \\ Y_{l,-l}(p_n), & \cdots & , Y_{l,l}(p_n) \end{bmatrix}.$$

We may tempted to directly estimate β in least squares fashion as $\widehat{\beta} = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{F}$. However, since the size of matrix $\mathbf{Y}'\mathbf{Y}$ becomes $(k+1)^2 \times (k+1)^2$, for large degree k, it may be difficult to directly invert the matrix. Instead of directly solving the normal equations, we project the normal equations into a smaller subspace $\mathcal{I}_l$ and estimate $2l+1$ coefficients in an iterative fashion.

At degree 0, we write $\mathbf{F} = \mathbf{Y}_0\beta_0 + \mathbf{r}_0$, where $\mathbf{r}_0$ is the residual vector of estimating $\mathbf{F}$ in the subspace $\mathcal{I}_0$. Note that the residual vector $\mathbf{r}_0$ consists of residuals $r_0(p_1), \cdots, r(p_n)$. Then we estimate β_0 by minimizing the residual vector in least squares fashion:

$$\widehat{\beta_0} = (\mathbf{Y}'_0\mathbf{Y}_0)^{-1}\mathbf{Y}'_0\mathbf{F} = \frac{\sum_{j=1}^n f(p_j)Y_{00}(p_j)}{\sum_{j=1}^n Y_{00}^2(p_j)}.$$

At degree l, we have

$$\mathbf{r}_{l-1} = e^{-l(l+1)\sigma}\mathbf{Y}_l\beta_l + \mathbf{r}_l, \tag{3.17}$$

where the residual vector $\mathbf{r}_{l-1}$ is obtained from the previous estimation as

$$\mathbf{r}_{l-1} = \mathbf{F} - \mathbf{Y}_0\widehat{\beta}_0 \cdots - e^{-(l-1)l\sigma}\mathbf{Y}_{l-1}\widehat{\beta}_{l-1}.$$

The least squares minimization of $\mathbf{r}_l$ is then given by

$$\widehat{\beta}_l = e^{l(l+1)\sigma}(\mathbf{Y}'_l\mathbf{Y}_l)^{-1}\mathbf{Y}'_l\mathbf{r}_{l-1}.$$

The IRF-algorithm is similar to the *matching pursuit method* although they were developed independently [82]. The IRF-algorithm was developed to avoid the computational burden of inverting a huge linear problem while the matching pursuit method was originally developed to compactly decompose a time frequency signal into a linear combination of pre-selected pool of basis functions called *dictionary*. In the usual least squares estimation with the design matrix Ψ of size $n \times k$ in (3.6), it is necessary to invert the $k \times k$ matrix $\Psi'\Psi$. Widely used matrix inversion algorithms such as Gauss-Jordan elimination, LU-decomposition and QR-decomposition, the running time is $\mathcal{O}(k^3) = \mathcal{O}(l^6)$, where $k = 1+3+\cdots+2l+1$ for using up to degree l [109, 66]. On the other hand, the IRF-algorithm, applied to spherical harmonics, requires to invert k number of submatrices of size $m \times m$ where $m = 1, 3, \cdots, 2l+1$. The total running time is then

$$\mathcal{O}(1^3 + 3^3 + \cdots + (2l+1)^3) = \mathcal{O}(l^4),$$

which is a substantial reduction of running time.

In the IRF-algorithm, we minimize the residual component $\mathbf{r}_l$ in least squares fashion, i.e. minimizing the sum of squared residuals $\sum_{j=1}^{n} r_l^2(\Omega_j)$ over all mesh vertices. On the other hand, in the marching pursuit method, the norm $\|\mathbf{Y}_l\beta_l\|^2$ is maximized. Due to orthonormality, maximizing the norm is equivalent to minimizing the norm of the residual $\|\mathbf{r}_l\|^2 = \int_{\mathcal{M}} r_l^2(p)\, d\mu(p)$. So there is a slight difference in how the residual is minimized. Although there is no limitation not to estimate multiple coefficients simultaneously in the matching pursuit method, [82] formulated it as the problem of estimating one coefficient at a time rather than multiple coefficients.

In many spherical harmonic representation literature [20, 53, 60, 106, 107], the optimal degree is simply selected based on a pre-specified error bound that depends on the size of anatomical structure. Although increasing the degree of the representation increases the goodness-of-fit, it also increases the number of coefficients to be estimated quadratically. So it is necessary to find the optimal degree where the goodness-of-fit and the number of parameters balance out. The stepwise model selection framework offers a way to automatically determine the optimal degree [31].

From (3.8), we can have

$$f(p_i) = \sum_{j=0}^{k-1} e^{-\lambda_j \sigma} \beta_j \psi_j(p_i) + \epsilon(p_i), \tag{3.18}$$

where $\epsilon(p_i)$ is a zero mean Gaussian random variable. Then we determine if adding the k-th degree terms in the $(k-1)$-th degree model (3.18) is statistically significant by testing the null hypothesis

$$H_0 : \mu_k = 0.$$

Let the k-th degree *sum of squared errors* (SSE) be

$$\mathrm{SSE}_k = \sum_{i=1}^{n} r_k^2(p_i).$$

As the degree k increases, SSE keep decreasing until it flattens out. So it is reasonable to stop the iteration when the decrease in error is no longer significant. Figure 3.1 shows the plot of the *root mean squared errors* (RMSE), $\sqrt{\mathrm{SSE}_k/n}$. Under H_0, the test statistic is

$$F = \frac{\mathrm{SSE}_{k-1} - \mathrm{SSE}_k}{\mathrm{SSE}_{k-1}/(n-k-1)} \sim F_{1,n-k-1},$$

the F-distribution with 1 and $n-k-1$ degrees of freedom. We compute the F statistic at each degree and stop the IRF procedure if the corresponding P-value first becomes bigger than the pre-specified significance α which is usually set at 0.05 (Figure 3.2).

3.4 Spherical Harmonics

The unit sphere S^2 can be parameterized by the polar angle θ and the azimuthal angel φ:

$$p = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) \tag{3.19}$$

with $p = (\theta, \varphi) \in [0, \pi] \otimes [0, 2\pi)$. The spherical Laplacian Δ corresponding to the parametrization (3.19) is then given by

$$\Delta = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\Big(\sin\theta\frac{\partial}{\partial\theta}\Big) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial^2\varphi}.$$

There are $2l+1$ eigenfunctions Y_{lm} ($-l \le m \le l$), corresponding to the same eigenvalue $\lambda_l = l(l+1)$ satisfying

$$\Delta Y_{lm} = \lambda_l Y_{lm}.$$

Y_{lm} is called the *spherical harmonic* of degree l and order m [37, 122]. It is given explicitly as

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

where $c_{lm} = \sqrt{\frac{2l+1}{2\pi}\frac{(l-|m|)!}{(l+|m|)!}}$ and P_l^m is the associated Legendre polynomials of order m. Unlike many previous imaging literatures on spherical harmonics that used the complex-valued spherical harmonics [20, 53, 60, 107], only real-valued spherical harmonics are used throughout the book for convenience in setting up a real-valued stochastic model.

For $f, h \in L^2(S^2)$, we define the inner product as

$$\langle f, h\rangle = \int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\pi} f(p)h(p)\, d\mu(p),$$

where Lebesgue measure $d\mu(p) = \sin\theta d\theta d\varphi$. Then with respect to the inner product, the spherical harmonics satisfies the orthonormal condition

$$\int_{S^2} Y_{ij}(p)Y_{lm}(p)\, d\mu(p) = \delta_{il}\delta_{jm},$$

where δ_{il} is the Kroneker's delta.

3.5 Spherical Harmonic Representation

The spherical harmonic (SPHARM) representation [18] has been applied to subcortical structures such as the hippocampus and the amygdala [53, 60, 72, 107]. In particular, [53] used the mean squared distance (MSD) of the SPHARM coefficients in quantifying ventricle surface shape in a twin study. [107] used the principal component analysis technique on the SPHARM coefficients of schizophrenic hippocampal surfaces in reducing the data dimension. Recently it has begun to be applied to more complex cortical surfaces [60, 106]. [60] presented SPHARM as a surface compression technique, where the main geometric features are encoded in the low degree spherical harmonics, while the noises are in the high degree spherical harmonics. In the SPHARM representation, all measurements are assigned equal weights and the coefficients of the series expansion is estimated in the least squares fashion. On the other hand, in the recently developed weighted version of SPHARM [31], closer measurements are weighted more and the coefficients of the series expansion is estimated in the weighted least squares fashion. So weighted-SPHARM is more suitable than SPHARM when the realization of the cortical boundaries, as triangle meshes, are noisy and possibly discontinuous.

In SPHARM, spherical harmonics are used in constructing the Fourier series expansion of the mapping from cortical surfaces to a unit sphere. So SPHARM is more of an interpolation technique than a smoothing technique, and thus it will have the ringing artifacts.On the other hand, the weighted version of SPHARM is a kernel smoothing technique given as a solution to a self-adjoint PDE [31]. The solution to the PDE is expanded in basis functions. In a similar spirit, [20] used the spherical harmonics in isotropic heat diffusion via the Fourier transform on a unit sphere as a form of hierarchical surface representation. The weighted-SPHARM offers many advantages over the previous PDE-based smoothing techniques [7, 34]. The PDE-based smoothing methods tend to suffer numerical instability while the weighted-SPHARM has no such problem [7, 22, 21, 34]. Since the traditional PDE-based smoothing gives an implicit numerical solution, setting up a statistical model is not straightforward. However, the weighted-SPHARM provides an explicit series expansion so it is easy to apply a wide variety of statistical modeling techniques such as the GLM [49] , principal component analysis (PCA) [107] and functional-PCA [87, 100]. The SPHARM-based global parametrization is computationally expensive compared to the local quadratic polynomial fitting [18, 34, 39, 71, 99] while providing more accuracy and flexibility for hierarchical representation.

Let $\mathcal{M}$ be a cortical surface topologically equivalent to a sphere. The unit sphere S^2 is realized as a triangle mesh and deformed to match the surface in such a way that anatomical homology and the topological connectivity of meshes are preserved. The cortical surfaces can be assumed to be smooth 2-dimensional Riemannian manifolds parameterized by two parameters [40, 71].

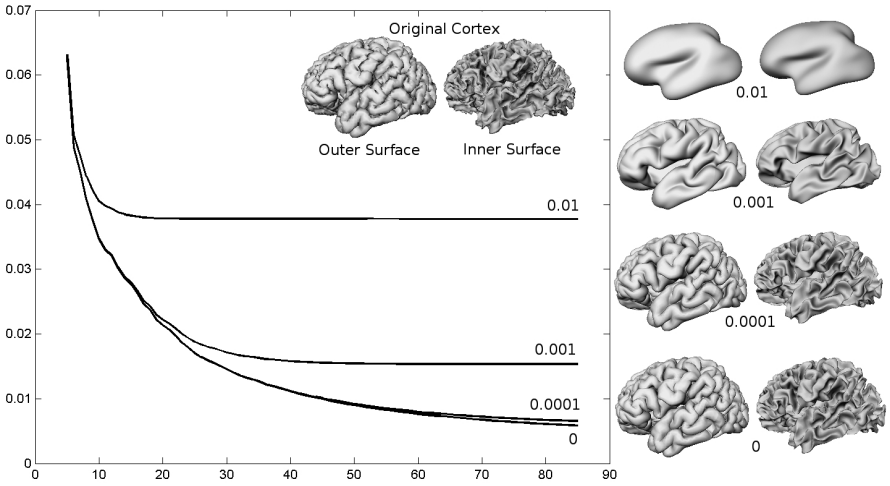

FIGURE 3.1
Plots of the RMSE for the weighted spherical harmonic representation with varying $\sigma(0.01, 0.001, 0.0001, 0)$. When $\sigma = 0$, we have the traditional spherical harmonic representation. The cortical surfaces correspond to the 85-th degree representation. As $\sigma \to 0$, the weighed representation converges to the traditional representation.

Based on the deformable algorithm [81] that establishes the homology between the S^2 mesh and the cortical surface, the Cartesian coordinates of the mapping are discretely parameterized by the spherical coordinates (3.19) as

$$v = (v_1(p), v_2(p), v_3(p)).$$

These discrete coordinate functions are further smoothed by the weighted-SPHARM:

$$v_i(p) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)\sigma} f_{lm}^i Y_{lm}(p). \tag{3.20}$$

Since SPHARM is a special case when $\sigma = 0$, we will simply present a model for the weighted-SPHARM only. We model v_i stochastically by assuming f_{lm}^i to follow independent normal distribution $N(\mu_{lm}^i, \sigma_l^2)$ for coordinate i, degree l, and order m. This assumption is equivalent to modeling v_i as the sum of signal plus noise:

$$v_i(p) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)t} \mu_{lm}^i Y_{lm}(p) + \epsilon_i(p),$$

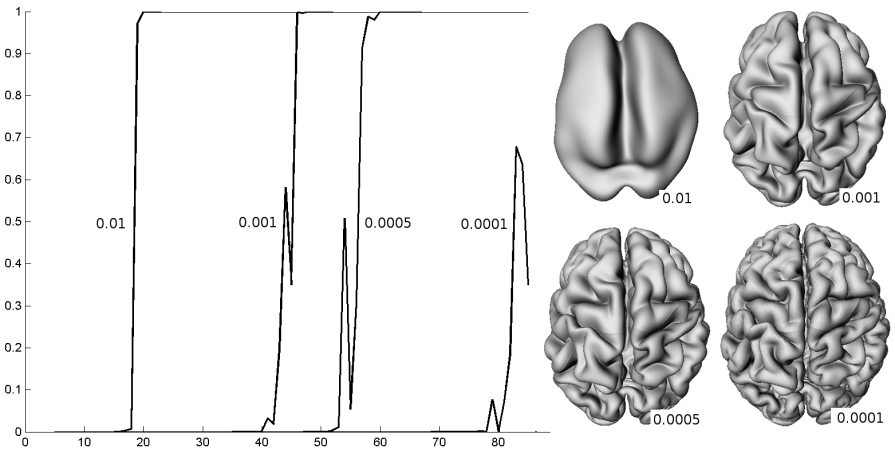

FIGURE 3.2
Cortical thickness projected onto the average outer cortex for various t and corresponding optimal degree: $k = 18(t = 0.01), k = 42(t = 0.001), k = 52(t = 0.0005), k = 78(t = 0.0001)$. The average cortex is constructed by averaging the coefficients of the weighted-SPHARM. The highly noise first image shows thickness measurements obtained by computing the distance between two triangle meshes.

where ϵ_i is a zero mean Guassian random field with a certain isotropic covariance function. A similar stochastic modeling approach has been used in [85], where the canonical expansion of a Gaussian random field is used to model the component of a deformation field.

The mean and the variance functions of the surface are then given by

$$\begin{aligned} \mathbb{E}v_i(p) &= \sum_{l=0}^{k}\sum_{m=-l}^{l} e^{-l(l+1)\sigma}\mu_{lm}^i Y_{lm}(p), \qquad (3.21)\\ \mathbb{V}v_i(p) &= \sum_{l=0}^{k}\sum_{m=-l}^{l} e^{-2l(l+1)\sigma}\sigma_l^2 Y_{lm}^2(p) \end{aligned}$$

The total variability of the surface is then measured by

$$\int_{S^2} \mathbb{V}v_i \, d\mu(p) = \sum_{l=0}^{k}\sum_{m=-l}^{l} e^{-2l(l+1)\sigma}\sigma_l^2$$

indicating the increase of smoothing bandwidth decreases the total variability. If

$$v_{ij}(\theta,\varphi) = \sum_{l=0}^{k}\sum_{m=-l}^{l} e^{-l(l+1)\sigma} f_{lm}^{ij} Y_{lm}(\theta,\varphi) \qquad (3.22)$$

is the weighted-SPHARM for the j-th subject ($1 \leq j \leq s$), the unknown parameters μ_{lm}^i and σ_l^2 are estimated as the sample mean and the sample variance:

$$\begin{aligned} \widehat{\mu_{lm}^i} &= \frac{1}{s}\sum_{j=1}^{n} f_{lm}^{ij}, \\ \widehat{\sigma_l^2} &= \frac{1}{(2l+1)(s-1)}\sum_{m=-l}^{l}\sum_{j=1}^{s}(f_{lm}^{ij}-\mu_{lm}^i)^2. \end{aligned} \tag{3.23}$$

For the numerical implementation, the iterative residual fitting (IRF) algorithm can be used. The `MATLAB` codes are available[1]. For up to $k = 78$ degree, there are total $3(k+1)^2 = 18,723$ unknown Fourier coefficients corresponding to the three Cartesian coordinates of a cortical surface.

The IRF-algorithm result can be validated against the analytical solution of equation. For any arbitrary initial data of the form

$$f = \sum_{l=0}^{k}\sum_{m=-l}^{l} \alpha_{lm} e^{l(l+1)\sigma} Y_{lm}, \tag{3.24}$$

the weighted-SPHARM representation is given by

$$K_\sigma * f = \sum_{l=0}^{k}\sum_{m=-l}^{l} \alpha_{lm} Y_{lm}. \tag{3.25}$$

Comparing the analytical expression (3.25) to the numerical result obtained from the IRF-algorithm serves as the basis for validation. It is sufficient to use a single term in (3.24) for validation. For the initial data $f = e^{l(l+1)\sigma}Y_{lm}$, we have $K_\sigma * f = Y_{lm}$. Table 3.5 shows the comparison for various degrees and orders. The fourth column shows the mean absolute error between the theoretical value Y_{lm} and the numerical result obtained from the IRF-algorithm. The mean is taken over all mesh vertices. As expected, the mean absolute error decreases as the degree increases. For the 78th degree with $\sigma = 0.0001$, the error is smaller than 2 decimal places. We have also checked if $\langle Y_{lm}, Y_{lm}\rangle$ is close to 1 in the last column. The estimation is accurate up to 2 decimal places for all degrees. Table 3.5 shows that the IRF-algorithm provides sufficiently good numerical accuracy.

3.6 Gibbs Phenomenon

The weakness of the traditional spherical harmonic representation is that it produces the Gibbs phenomenon (ringing artifacts) for discontinuous and

[1] `http://www.stat.wisc.edu/~mchung/softwares/weighted-SPHARM/weighted-SPHARM.html`

The numerical accuracy of the weighted-SPHARM representation against the analytic solution f_{lm} for various bandwidth σ.

degree l	order m	bandwidth σ	mean absolute error	$\langle Y_{lm}, Y_{lm} \rangle$
18	17	0	0.0077	0.9979
18	17	0.0001	0.0078	0.9979
18	17	0.0005	0.0083	0.9981
18	17	0.01	0.0575	0.9995
42	41	0	0.0064	0.9977
42	41	0.001	0.0126	0.9992
52	51	0	0.0066	0.9972
52	51	0.0005	0.0101	0.9988
78	77	0	0.0060	0.9973
78	77	0.0001	0.0068	0.9984

rapidly changing continuous measurements [51, 31]. Gibbs phenomenon often arises in Fourier series expansion of discontinuous data. It is named after American physicist Josiah Willard Gibbs. In representing a piecewise continuously differentiable data using the Fourier series, the overshoot of the series happens at a jump discontinuity (Figure 3.3). The overshoot does not decease as the number of terms increases in the series expansion, and it converges to a finite limit called the Gibbs constant.

The Gibbs phenomenon was first observed by Henry Willbraham in 1848 but it did not attract any attention at that time [123]. Then a Nobel prize laureate Albert Michelson constructed an harmonic analyzer, one of the first mechanical analogue computers, that was used to plots Fourier series and observed the phenomenon. He thought the phenomenon was caused by mechanical error. Josiah W. Gibbs rediscovered the phenomenon in 1898 [54] and correctly explained the phenomenon as mathematical in 1899 [55]. Later mathematician Maxime Bocher named it the Gibbs phenomenon and gave a precise mathematical analysis in 1906 [16]. The Gibbs phenomenon associated with spherical harmonics were first observed by Herman Weyl in 1968 [**?**]. The history and the overview of Gibbs phenomenon can be found in several literature [48, 67].

Consider the finite Fourier series expansion of 1D piecewise smooth function $f \geq 0$ with discontinuity at c given by

$$S_k(u) = \sum_{j=0}^{k} f_j \psi_j(u),$$

where $f_j = \langle f, \psi_j \rangle$. The basis is the usual sin and cosine functions. Let

$$d = \lim_{u \to c^+} f(u) - \lim_{u \to c^-} f(u) > 0$$

be the size of jump. Let $u_\circ$ be the first local maximum. Then the amount of

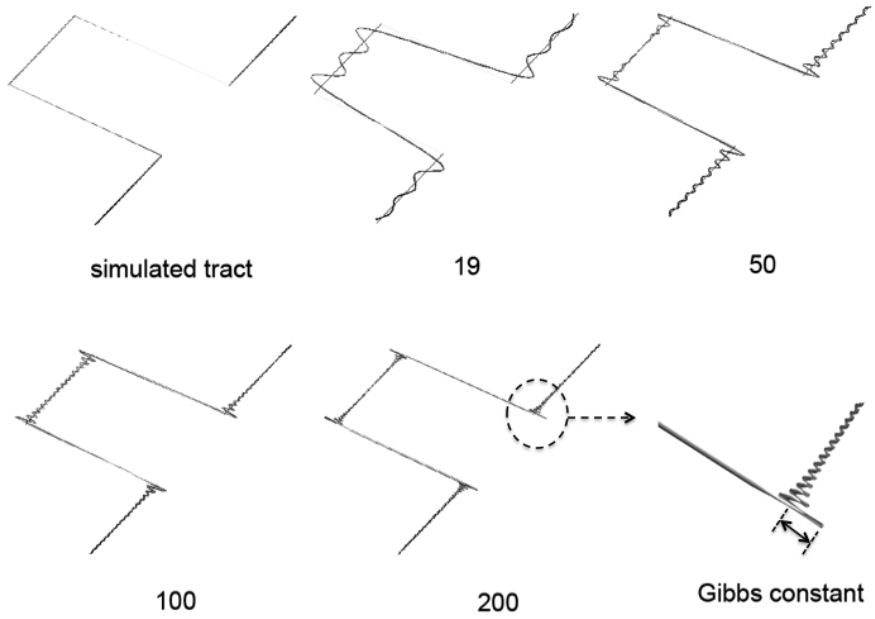

FIGURE 3.3
Gibbs phenomenon on the cosine series representation of a simulated tract for degrees 19, 50, 100 and 200 [29]. Increasing the number of basis does not reduce the overshoot at the corner. The maximum overshoot is proportional to Gibbs constant.

overshoot associated with the k-th series expansion is given by

$$S_k(u_\circ) - \lim_{u \to c^+} f(u).$$

Then we can show that the limit of the overshoot is

$$\lim_{k \to \infty} S_k(u_\circ) - \lim_{u \to c^+} f(u) = \frac{d}{2}(g-1),$$

where the Gibbs constant g is given by

$$g = \frac{2}{\pi} \int_0^\pi \frac{\sin x}{x}\, dx = 1.17897974 \cdots .$$

Figure 3.3 shows the overshoot in the cosine series representation [29]. In Figure 3.3, we have simulated 300 uniformly sampled control points along the parameterized curve $(x, y, z) = (t, 0, t)$ for $t \in [1, 100) \cup [200, 300)$ and

$(x, y, z) = (t, 1, t)$ for $t \in [200, 300)$. The control points are fitted with the cosine representation with various degrees. As the degree increases to 200, the representation suffers from the severe ringing artifacts. The overshoot does not disappear even as the degree of expansion goes to infinity. Various neuroanatomical curvilinear structures are such as white matter fibers and corpus callosum boundaries are supposed to be smooth so we will not likely to encounter the severe Gibbs phenomenon.

There are few available techniques for reducing Gibbs phenomenon [19, 57]. Most techniques are variation on some sort of kernel methods. One of the standard method is to use the Fejer kernel which is defined as

$$K_n(u) = \frac{1}{n} \sum_{j=0}^{n-1} D_j(u),$$

where D_j is the Dirichlet kernel

$$D_j = \sum_{k=-j}^{j} e^{iku}.$$

Then it can be shown that

$$K_n(u) = \frac{1}{n} \Big(\frac{\sin \frac{nu}{2}}{\sin \frac{u}{2}} \Big)^2.$$

The kernel is symmetric and positive. Then we have

$$K_n * f \to f$$

for any $f \in L^2([-\pi, \pi])$ as $n \to \infty$. Since kernel is unimodal, it has the effect of smoothing the discontinuous signal f and in turn the convolution will not exhibit the ringing artifacts for sufficiently large n. Heat kernel smoothing and weighted Fourier representation behave similarly and can be used in reducing Gibbs phenomenon.

The Gibbs phenomenon will likely arise in modeling arbitrary anatomical objects with possible sharp corners. The Gibbs phenomenon can be effectively removed if the spherical harmonic representation converges faster as the degree goes to infinity. By weighting the spherical harmonic coefficients exponentially smaller, we can make the representation converges faster. This can be achieved by additionally weighting the spherical harmonic coefficients with the heat kernel. Figure 3.4 demonstrates the severe Gibbs phenomenon in the traditional spherical harmonic representation (top) on a hat shaped 2D surface. The hat shaped surface is simulated as $z = 1$ for $x^2 + y^2 < 1$ and $z = 0$ for $1 \leq x^2 + y^2 \leq 2$. On the other hand the weighted spherical harmonic representation (bottom) shows substantially reduced ringing artifacts. Due to very complex folding patterns, sulcal regions of the brain exhibit more abrupt

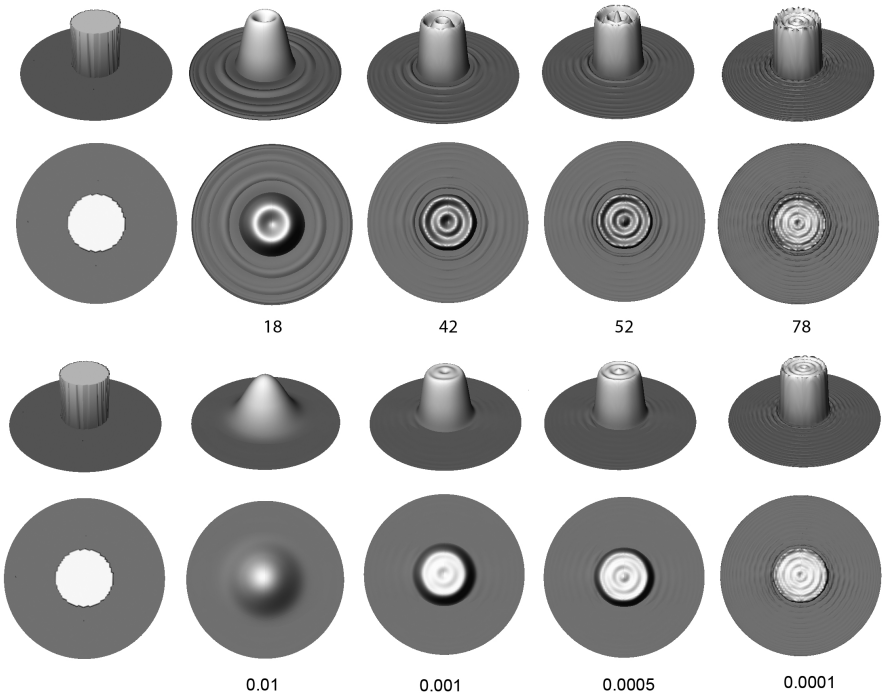

FIGURE 3.4
Gibbs phenomenon on a hat shaped simulated surface. The SPHARM representation (top) of degrees 18, 42, 52 and 78 show severe ringing artifacts. One the other hand, the weighted-SPHARM representation (bottom) with bandwidths 0.01, 0.001, 0.0005, 0.0001 shows less ringing artifacts. The optimal degrees for the weighed representation is determined by the model selection procedure and found to be 18, 42, 52 and 78 respectively.

directional change than the simulated hat surface(upward of 180 degree compared to 90 degree in the hat surface) so there is a need for reducing the Gibbs phenomenon in the traditional spherical harmonic representation.

In Figure 3.5, a different example is given for Gibbs phenomenon. Discontinuous measurements are constructed as a step function of value 1 in the circular band $\frac{1}{8} < \theta < \frac{1}{4}$ and 0 outside of the band on a unit sphere. The SPAHRM representation of the step function resulted in significant ringing artifacts even for fairly high degrees up to $k = 78$. In comparison, the weighted-SPHARM representation does not exhibit any serious ringing artifacts. The superior performance of the weighted-SPHARM can be easily explained in terms of convergence. The weighted-SPHARM representation additionally weights Fourier coefficients with exponentially decaying weights,

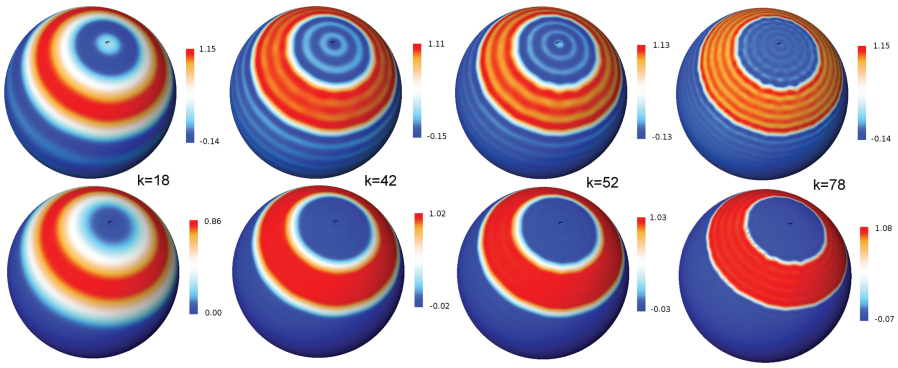

FIGURE 3.5
Gibbs phenomenon in the SPHARM representation for degrees 18, 42, 52 and 78. The traditional SPHARM (top) and the weighted representations (bottom) are performed on the discontinuous measurements on a unit sphere, which are defined as 1 in region $\frac{1}{8} < \theta < \frac{1}{4}$ and 0 in other regions. The SPHARM representation shows severe ringing artifacts while the weighted-SPHARM shows the negligible ringing effect.

which contributes more rapid convergence even for discontinuous measurements. This robustness of weighted-SPHARM is also related to the fact that it is a PDE-based data smoothing technique while the traditional SPHARM is more of interpolation or reconstruction technique.

The amount of the Gibbs phenomenon can be numerically quantified using the *overshoot* as the maximum of L_2 norm of the residual difference between the original and the reconstructed surface as

$$\sup_{(\theta,\varphi)\in S^2} \left|\left| p(\theta,\varphi) - \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)\sigma} f_{lm} Y_{lm}(\theta,\varphi) \right|\right|.$$

If surface coordinates are abruptly changing or their derivatives are discontinuous, the Gibbs phenomenon will severely distort the surface shape and the overshoot will never converge to zero. We have reconstructed a cube with various degree presentation and the bandwidth showing more ringing artifacts and overshoot in the traditional representation compared to the proposed weighted version (Figure 3.6). The exponentially decaying weights make the representation converge faster and reduce the Gibbs phenomenon significantly. The plots in Figure 3.6 display the amount of overshoot for the traditional representation (black) and the weighted version (red). The weighted spherical harmonic representation shows smaller overshoot compared to the traditional representation.

For the random field-based inference on a sphere, it is useful to properly

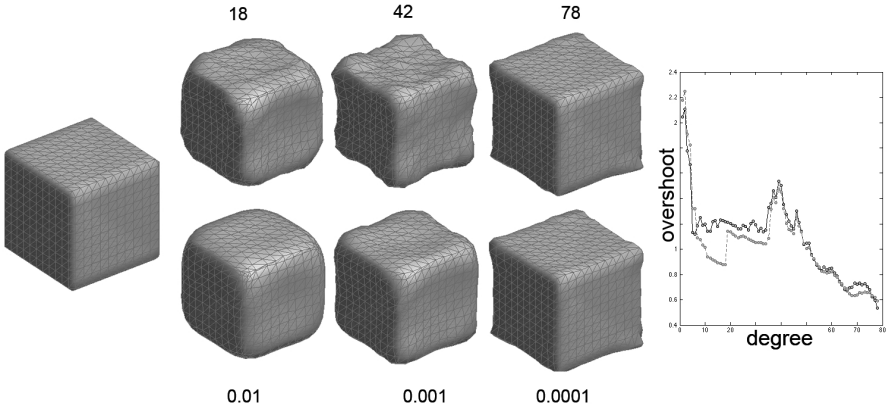

FIGURE 3.6
The Gibbs phenomenon in the spherical harmonic representation (top) of a cube for degrees $k = 18, 42, 78$. The weighted spherical harmonic representation (bottom) at the same degrees but with bandwidth $\sigma = 0.01, 0.001, 0.0001$ respectively. The plots display the amount of overshoot for the traditional representation (black) vs. the weighted version (grey). In almost all degrees, the traditional spherical harmonic representation shows more prominent Gibbs phenomenon compared to the weighted version.

estimate the full width at the half maximum (FWHM) of a heat kernel [32, 132] . Estimating the FWHM of the heat kernel is not trivial since there is no known close form expression for FWHM as a function of bandwidth σ in a curved surface. So it is necessary to estimate FWHM is numerically. For $p, q, r \in S^2$, let us define the Cartesian inner product $\cdot$ as $p \cdot q = \cos(\theta)$, where θ is an angle between p and q. The heat kernel is symmetric along the geodesic circle. If $p \cdot q = p \cdot r$, we have $K_\sigma(p, q) = K_\sigma(p, r)$. This property can be used to simplify the expansion using the harmonic addition theorem [59, 122], which states that

$$\sum_{m=-l}^{l} Y_{lm}(p) Y_{lm}(q) = \frac{2l+1}{4\pi} P_l^0(p \cdot q). \tag{3.26}$$

For any $p, q \in S^2$, the heat kernel can now be simplified as

$$K_\sigma(p, q) = \sum_{l=0}^{k} \frac{2l+1}{4\pi} e^{-l(l+1)\sigma} P_l^0(p \cdot q). \tag{3.27}$$

The expression (3.27) is used to plot the shape of the heat kernel by fixing p to be the north pole and by varying $\theta = \cos^{-1}(p \cdot q)$. A similar result is

also given in [20]. The maximum of the kernel is obtained at $\theta = 0$. Then the FWHM is solved numerically for θ in the equation

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

The FWHM is then 2θ.

3.7 Surface Registration

Previously cortical surface normalization was performed by minimizing an objective function that measures the global fit of two surfaces while maximizing the smoothness of the deformation in such a way that the gyral patterns are matched smoothly [32, 102, 118]. In the spherical harmonic representation, the surface normalization is straightforward and does not require any sort of optimization explicitly but at least requires some initial alignment. A crude alignment can be done by coinciding the first order ellipsoid meridian and equator in the SPHARM-correspondence approach [53, 111]. For cortical meshes obtained using the anatomic segmentation using the proximities (ASP) algorithm [81], such alignments are not needed. An approximate surface alignment is done during the cortical surface extraction process. The algorithm generates 40,962 vertices and 81,920 triangles with the identical mesh topology for all subjects. The vertices indexed identically on two cortical meshes will have a very close anatomic homology and this defines the surface alignment. This provides the same spherical parameterization at identically indexed vertices across different cortical surfaces.

Consider a surface $h = (h_1, h_2, h_3)$ obtained from the coordinates v_i measured at point p:

$$h_i(p) = \sum_{l=0}^{k}\sum_{m=-l}^{l}\langle v_i, Y_{lm}\rangle(p).$$

Consider another surface j_i obtained from coordinate functions w_i:

$$j_i(p) = \sum_{l=0}^{k}\sum_{m=-l}^{l}\langle w_i, Y_{lm}\rangle(p).$$

Suppose the surface h_i is deformed to $h_i + d_i$ under the influence of the displacement vector field d_i. We wish to find d_i that minimizes the discrepancy between $h_i + d_i$ and j_i in the finite subspace $\mathcal{H}_k$. This can be easily done by noting that

$$\sum_{l=0}^{k}\sum_{m=-l}^{l}(w_{lm}^i - v_{lm}^i)Y_{lm}(p) = \arg\min_{d_i\in\mathcal{H}_k}\left|\left|\widehat{h_i} + d_i - \widehat{j_i}\right|\right|. \tag{3.28}$$

This implies that the optimal displacement in the least squares sense is obtained by simply taking the difference between two weighted spherical harmonic representation and matching coefficients of the same degree and order. Then a specific point $\widehat{h_i}(p_0)$ in one surface corresponds to $\widehat{j_i}(p_0)$ in the other surface. We refer to this point-to-point surface correspondence as the *spherical harmonic correspondence* [31]. The spherical harmonic correspondence shows that the optimal displacement in the least squares sense is obtained by simply taking the difference between two spherical harmonic representations. Unlike other surface registration methods used in warping surfaces between subjects [32, 102, 118], it is not necessary to consider an additional cost function that guarantees the smoothness of the displacement field since the displacement field $d = (d_1, d_2, d_3)$ is already a linear combination of smooth basis functions.

The previously available approaches for computing the cortical thickness in discrete triangle meshes produce noisy thickness measures [32, 46, 81]. So it is necessary to smooth the thickness measurements along the cortex via surface-based smoothing techniques [7, 22, 21, 34]. On the other hand, the weighted-SPHARM provides smooth functional representation of the outer and inner surfaces so that the distance measures between the surfaces should be already smooth.Hence, the weighted-SPHARM avoids the additional step of thickness smoothing done in most of thickness analysis literature [32, 34] while it is not necessary to perform data smoothing in the spherical harmonic formulation. The distance between the outer and inner cortical surfaces can be determined using the spherical harmonic correspondence. Given the outer surface h_i and the inner surface j_i, the cortical thickness is defined to be the Euclidean distance between the two representations:

$$\texttt{thick}(p) = \sqrt{\sum_{i=1}^{3} \Big[\sum_{l=0}^{k} \sum_{m=-l}^{l} \langle v_i - w_i, Y_{lm} \rangle \Big]^2}.$$

A similar approach has been proposed for measuring the closeness between two surfaces [53]. Figure 3.7 shows the comparison of cortical thickness computed from the traditional deformable surface algorithm [81] and the spherical harmonic correspondence. The cortical thickness obtained from the traditional approach introduces a lot of triangle mesh noise into its estimation while the spherical harmonic correspondence approach does not. The spatial smoothness of the thickness is explicitly incorporated via the bandwidth σ.

3.8 Cortical Asymmetry

Previous neuroanatomical studies have shown left occipital and rigtht frontal lobe asymmetry, and left planum temporal asymmetry in normal controls [12, 73, 120]. These studies mainly flip the whole brain 3D MRI to obtain the

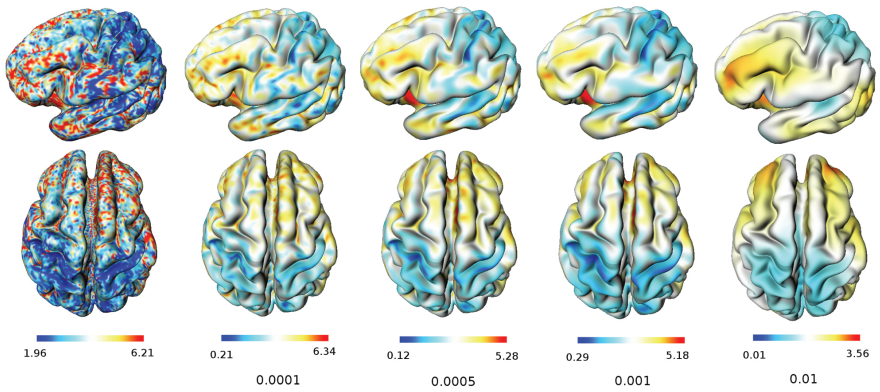

FIGURE 3.7
Cortical thickness of a subject projected onto a template. The cortical thickness is computed from the spherical harmonic correspondence with heat kernel weights. As the bandwidth increases from $\sigma = 0.0001$ to 0.01, the amount of smoothing also increases. The first image shows the cortical thickness obtained from the traditional deformable surface algorithm [81].

mirror reflected MRI with respect to the mid-saggital cross-section. Then the anatomical correspondence across the hemispheres is established and a subsequent statistical analysis is performed at each voxel in the 3D MRI. Although this approach is sufficient for the voxel-based morphometry [8], where we only need an approximate alignment of corresponding brain substructures, it may fail to properly align highly convoluted sulcal and gyral foldings of gray matter. In order to address this shortcoming inherent in 3D whole brain volume asymmetry analysis, we need a new 2D cortical surface based framework.

As shown in the previous section, surface correspondence between two surfaces can be established using spherical harmonics. For asymmetry, we also need to establish hemispheric correspondence within a subject. However, it is not straightforward to establish a 2D surface-based hemispheric correspondence. Although there are many 3D volume-based brain hemisphere asymmetry analyses [12, 73], due to this simple reason, there is a lack of 2D surface-based asymmetry analyses. This will be the first unified mathematical framework on 2D cortical asymmetry. The inherent angular symmetry presented in the weighted spherical harmonic representation can be used to establish the inter-hemispheric correspondence. It turns out that the usual asymmetry index of (L-R)/(L+R) is expressed as the ratio between the sum of positive and negative order harmonics.

The spherical harmonic correspondence described in the previous section can be further used to establish the inter-hemispheric correspondence by letting $\widehat{j}_i$ be the mirror reflection of $\widehat{h}_i$. The mirror reflection of $\widehat{h}_i$

with respect to the midsaggital cross section $u_2 = 0$ is simply given by $\widehat{j}_i(\theta,\varphi) = \widehat{h}_i{}^*(\theta,\varphi) = \widehat{h}_i(\theta, 2\pi - \varphi)$, where * denotes the mirror reflection operation. The specific point $\widehat{h}_i(\theta_0,\varphi_0)$ in the left hemisphere will be mirror reflected to $\widehat{j}_i(\theta_0, 2\pi-\varphi_0)$ in the right hemisphere. The spherical harmonic correspondence of $\widehat{j}_i(\theta_0, 2\pi-\varphi_0)$ is $\widehat{h}_i(\theta_0, 2\pi-\varphi_0)$. Hence, the point $\widehat{h}_i(\theta_0,\varphi_0)$ in the left hemisphere corresponds to the point $\widehat{h}_i(\theta_0, 2\pi-\varphi_0)$ in the right hemisphere. This establishes the inter-hemispheric anatomical correspondence.

The inter-hemispheric correspondence is used to compare cortical thickness measurements f across the hemispheres. The weighted spherical harmonic representation of cortical thickness f is

$$\widehat{g}(\theta,\varphi) = \sum_{l=0}^{k}\sum_{m=-l}^{l} e^{-l(l+1)\sigma}\langle f, Y_{lm}\rangle Y_{lm}(\theta,\varphi).$$

At a given position $h_i(\theta_0,\varphi_0)$, the corresponding cortical thickness is $\widehat{g}(\theta_0,\varphi_0)$, which should be compared with the thickness $\widehat{g}(\theta_0, 2\pi - \varphi_0)$ at position $\widehat{h}_i(\theta_0, 2\pi - \varphi_0)$:

$$\widehat{g}(\theta_0, 2\pi-\varphi_0) = \sum_{l=0}^{k}\sum_{m=-l}^{l} e^{-l(l+1)\sigma}\langle f, Y_{lm}\rangle Y_{lm}(\theta, 2\pi-\varphi). \tag{3.29}$$

The equation (3.29) can be rewritten using the property of spherical harmonics:

$$Y_{lm}(\theta, 2\pi-\varphi) = \begin{cases} -Y_{lm}(\theta,\varphi), & -l \le m \le -1, \\ Y_{lm}(\theta,\varphi), & 0 \le m \le l, \end{cases}$$

$$\begin{aligned}\widehat{g}(\theta_0, 2\pi-\varphi_0) &= \sum_{l=0}^{k}\sum_{m=-l}^{-1} e^{-l(l+1)\sigma}\langle f, Y_{lm}\rangle Y_{lm}(\theta_0,\varphi_0) \\ &\quad - \sum_{l=0}^{k}\sum_{m=0}^{l} e^{-l(l+1)\sigma}\langle f, Y_{lm}\rangle Y_{lm}(\theta_0,\varphi_0).\end{aligned}$$

Comparing with the expansion for $\widehat{g}(\theta_0,\varphi_0)$, we see that the negative order terms are invariant while the positive order terms change sign. Hence we define the *symmetry index* as

$$S(\theta,\varphi) = \frac{1}{2}\Big[\widehat{g}(\theta,\varphi) + \widehat{g}(\theta, 2\pi-\varphi)\Big] = \sum_{l=0}^{k}\sum_{m=-l}^{-1} e^{-l(l+1)\sigma}\langle f, Y_{lm}\rangle Y_{lm}(\theta_0,\varphi_0),$$

and the *asymmetry index* as

$$A(\theta,\varphi) = \frac{1}{2}\Big[\widehat{g}(\theta,\varphi) - \widehat{g}(\theta, 2\pi-\varphi)\Big] = \sum_{l=0}^{k}\sum_{m=0}^{l} e^{-l(l+1)\sigma}\langle f, Y_{lm}\rangle Y_{lm}(\theta_0,\varphi_0).$$

We normalize the asymmetry index by dividing it by the symmetry index as

$$\begin{aligned} N(\theta,\varphi) &= \frac{\widehat{g}(\theta,\varphi)-\widehat{g}(\theta,2\pi-\varphi)}{\widehat{g}(\theta,\varphi)+\widehat{g}(\theta,2\pi-\varphi)} \\ &= \frac{\sum_{l=1}^{k}\sum_{m=-l}^{-1} e^{-1(l+1)\sigma}\langle f, Y_{lm}\rangle Y_{lm}(\theta,\varphi)}{\sum_{l=0}^{k}\sum_{m=0}^{l} e^{-l(l+1)\sigma}\langle f, Y_{lm}\rangle Y_{lm}(\theta,\varphi)}. \end{aligned}$$

We refer to this index as the *normalized asymmetry index*. The numerator is the sum of all negative orders while the denominator is the sum of all positive and the 0-th orders. Note that $N(\theta,0) = N(\theta,\pi) = 0$. This index is intuitively interpreted as the normalized difference between cortical thickness in the left and the right hemispheres. Note that the larger the value of the index, the larger the amount of asymmetry. The index is invariant under the affine scaling of the human brain so it is not necessary to control for the global brain size difference in the later statistical analysis.

For each subject, its normalized asymmetry index $N(\theta,\varphi)$ is computed and modeled as a zero mean Gaussian random field. The null hypothesis is that $N(\theta,\varphi)$ is identical in the both groups for all (θ,φ), while the alternate hypothesis is that there is a specific point (θ_0,φ_0) at which the normalized asymmetry index is different. For the traditional group comparison between autistic and normal control subjects, the T statistic at each point (θ,φ) would be constructed. Since T statistics at different points are correlated, it becomes a *multiple comparison problem* [89, 133]. The corrected P-value accounting for spatially correlated test statistics is determined by computing the distribution of the supremum of T random field [133], i.e.

$$P\Big[\sup_{(\theta,\varphi)\in S^2} T(\theta,\varphi) < h\Big]. \tag{3.30}$$

Bibliography

[1] R.J. Adler. *The Geometry of Random Fields*. John Wiley & Sons, 1981.

[2] R.J. Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. IMS, Hayward, CA, 1990.

[3] R.J. Adler. On excursion sets, tube formulas and maxima of random fields. *Annals of Applied Probability*, 10:1–74, 2000.

[4] R.J. Adler, G. Samorodnitsky, and T. Gadrich. The expected number of level crossings for stationary, harmonisable, symmetric, stable processes. *The Annals of Applied Probability*, 3:553–575, 1993.

[5] R.J. Adler and J.E. Taylor. *Random fields and geometry*. Springer Verlag, 2007.

[6] D. Aldous. *Probability Approximations via the Poisson Clumping Heuristic*. Springer-Verlag, New York, 1989.

[7] A. Andrade, F Kherif, J. Mangin, K.J. Worsley, A. Paradis, O. Simon, S. Dehaene, D. Le Bihan, and J-B. Poline. Detection of fmri activation using cortical surface mapping. *Human Brain Mapping*, 12:79–93, 2001.

[8] J. Ashburner and K. Friston. Voxel-based morphometry - the methods. *NeuroImage*, 11:805–821, 2000.

[9] J. Ashburner, C. Hutton, R. S. J. Frackowiak, I. Johnsrude, C. Price, and K. J. Friston. Identifying global anatomical differences: deformation-based morphometry. *Human Brain Mapping*, 6:348–357, 1998.

[10] J. Ashburner, P. Neelin, D.L. Collins, A.C. Evans, and K.J. Friston. Incorporating prior knowledge into image registration. *NeuroImage*, 6:344–352, 1997.

[11] E.H. Aylward, N.J Minshew, G. Goldstein, N.A. Honeycutt, A.M. Augustine, K.O. Yates, P.E. Bartra, and G.D. Pearlson. Mri volumes of amygdala and hippocampus in nonmentally retarded autistic adolescents and adults. *Neurology*, 53:2145–2150, 1999.

[12] T.R. Barrick, C.E. Mackay, S. Prima, F. maes, D. Vandermeulen, T.J. Crow, and N. Roberts. Automatic analysis of cerebral asymmetry: an exploratory study of the relationship between brain torgue and planum temporale asymmetry. *NeuroImage*, 24:678–691, 2005.

[13] I.V. Basawa and B.L.S.P. Rao. *Statistical Inference for Stochastic Processes*. Academic Press, London, 1980.

[14] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc, Ser. B*, 57:289–300, 1995.

[15] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001.

[16] M. Bocher. Introduction to the theory of Fourier's series. *Ann. Math*, 7:81–152, 1906.

[17] P. Brambilla, M.A. Nicoletti, R.B. Sassi, A.G. Mallinger, E. Frank, D.J. Kupfer, M.S. Keshavan, and J.C. Soares. Magnetic resonance imaging study of corpus callosum abnormalities in patients with bipolar disorder. *Biological psychiatry*, 54:1294–1297, 2003.

[18] C. Brechbuhler, G. Gerig, and O. Kubler. Parametrization of closed surfaces for 3d shape description. *Computer Vision and Image Understanding*, 61:154–170, 1995.

[19] C. Brezinski. Extrapolation algorithms for filtering series of functions, and treating the Gibbs phenomenon. *Numerical Algorithms*, 36:309–329, 2004.

[20] T. Bulow. Spherical diffusion for 3D surface smoothing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1650–1654, 2004.

[21] A. Cachia, J.-F. Mangin, D. Riviere, F. Kherif, N. Boddaert, A. Andrade, D. Papadopoulos-Orfanos, J.-B. Poline, I. Bloch, M. Zilbovicius, P. Sonigo, F. Brunelle, and J. . Regis. A primal sketch of the cortex mean curvature: a morphogenesis based approach to study the variability of the folding patterns. *IEEE Transactions on Medical Imaging*, 22:754–765, 2003.

[22] A. Cachia, J.-F. Mangin, D. Riviére, D. Papadopoulos-Orfanos, F. Kherif, I. Bloch, and J. Régis. A generic framework for parcellation of the cortical surface into gyri using geodesic voronoï diagrams. *Image Analysis*, 7:403–416, 2003.

[23] J. Cao. The size of the connected components of excursion sets of χ^2, t and F fields. *Advances in Applied Probability*, 31:579–595, 1999.

[24] J. Cao and K. Worsley. The geometry of correlation fields with an application to functional connectivity of the brain. *Annals of Applied Probability*, 9:1021–1057, 1999.

[25] J. Cao and K. J. Worsley. The detection of local shape changes via the geometry of Hotellings t2 fields. *Annals of Statistics*, 27:925–942, 1999.

[26] J. Cao and KJ Worsley. Applications of random fields in human brain mapping. *Spatial Statistics: Methodological Aspects and Applications*, 159:170–182, 2001.

[27] J. Cates, P.T. Fletcher, M. Styner, H.C. Hazlett, and R. Whitaker. Particle-Based Shape Analysis of Multi-Object Complexes. In *Medical image computing and computer-assisted intervention: MICCAI... International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 11, pages 477–485, 2008.

[28] M.K. Chung. Heat kernel smoothing on unit sphere. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, volume I, pages 467–473, 2006.

[29] M.K. Chung, N. Adluru, J.E. Lee, M. Lazar, J.E. Lainhart, and A.L. Alexander. Cosine series representation of 3d curves and its application to white matter fiber bundles in diffusion tensor imaging. *Statistics and Its Interface*, 3:69–80, 2010.

[30] M.K. Chung, K.M. Dalton, A.L. Alexander, and R.J. Davidson. Less white matter concentration in autism: 2D voxel-based morphometry. *NeuroImage*, 23:242–251, 2004.

[31] M.K. Chung, K.M. Dalton, L. Shen, A.C. Evans, and R.J. Davidson. Weighted Fourier representation and its application to quantifying the amount of gray matter. *IEEE Transactions on Medical Imaging*, 26:566–581, 2007.

[32] M.K. Chung, S. Robbins, Davidson R.J. Alexander A.L. Dalton, K.M., and A.C. Evans. Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, 25:1256–1265, 2005.

[33] M.K. Chung, K.J. Worsley, T. Paus, D.L. Cherif, C. Collins, J. Giedd, J.L. Rapoport, , and A.C. Evans. A unified statistical approach to deformation-based morphometry. *NeuroImage*, 14:595–606, 2001.

[34] M.K. Chung, K.J. Worsley, S. Robbins, T. Paus, J. Taylor, J.N. Giedd, J.L. Rapoport, and A.C. Evans. Deformation-based surface morphometry applied to gray matter deformation. *NeuroImage*, 18:198–213, 2003.

[35] M.K. Chung, K.J. Worsley, J. Taylor, J. Ramsay, S. Robbins, and A.C. Evans. Diffusion smoothing on the cortical surface. *NeuroImage*, 13(6S1):95, 2001.

[36] D. L. Collins, T. Paus, A. Zijdenbos, K. J. Worsley, J. Blumenthal, J. N. Giedd, J. L. Rapoport, and A. C. Evans. Age related changes in

the shape of temporal and frontal lobes: An mri study of children and adolescents. *Soc. Neurosci. Abstr.*, 24:304, 1998.

[37] R. Courant and D. Hilbert. *Methods of Mathematical Physics*. Interscience, New York, english edition, 1953.

[38] J.G. Csernansky, L. Wang, S.C. Joshi, J. Tilak Ratnanather, and M.I. Miller. Computational anatomy and neuropsychiatric disease: probabilistic assessment of variation and statistical inference of group difference, hemispheric asymmetry, and time-dependent change. *NeuroImage*, 23:56–68, 2004.

[39] A.M. Dale and B. Fischl. Cortical surface-based analysis i. segmentation and surface reconstruction. *NeuroImage*, 9:179–194, 1999.

[40] C. Davatzikos and R.N. Bryan. Using a deformable surface model to obtain a shape representation of the cortex. *Proceedings of the IEEE International Conference on Computer Vision*, 9:2122–2127, 1995.

[41] A.P. Dempster, N.M. Laird, D.B. Rubin, et al. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:1–38, 1977.

[42] E.R. Dougherty. *Random Processes for Image and Signal Processing*. IEEE Press, 1999.

[43] B. Egaas, E. Courchesne, and O. Saitoh. Reduced size of corpus callosum in autism. *Archives of Neurology*, 52:794, 1995.

[44] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC, 1996.

[45] W. Feller. Introduction to probability theory and its applications, Vol. 1. 1968.

[46] B. Fischl and A.M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *PNAS*, 97:11050–11055, 2000.

[47] B. Flury. *A First Course in Multivariate Statistics*. Springer, 1997.

[48] J. Foster and FB Richards. The Gibbs phenomenon for piecewise-linear approximation. *American Mathematical Monthly*, 98, 1991.

[49] K.J. Friston. A short history of statistical parametric mapping in functional neuroimaging. Technical Report Technical report, Wellcome Department of Imaging Neuroscience, ION, UCL., London, UK., 2002.

[50] C. Gaser, H.-P. Volz, S. Kiebel, S. Riehemann, and H. Sauer. Detecting structural changes in whole brain based on nonlinear deformationsapplication to schizophrenia research. *NeuroImage*, 10:107–113, 1999.

[51] A. Gelb. The resolution of the gibbs phenomenon for spherical harmonics. *Mathematics of Computation*, 66:699–717, 1997.

[52] C.R. Genovese, N.A. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15:870–878, 2002.

[53] G. Gerig, M. Styner, D. Jones, D. Weinberger, and J. Lieberman. Shape analysis of brain ventricles using spharm. In *MMBIA*, pages 171–178, 2001.

[54] J.W. Gibbs. Fouriers series. *Nature*, 59:200, 1898.

[55] J.W. Gibbs. Fourier's series. *Nature*, 59:606, 1899.

[56] P. Golland, W.E.L. Grimson, M.E. Shenton, and R. Kikinis. Deformation analysis for shape based classification. *Lecture Notes in Computer Science*, pages 517–530, 2001.

[57] D. Gottlieb and C.W. Shu. On the Gibbs phenomenon and its resolution. *SIAM Review*, pages 644–668, 1997.

[58] H. Gray. *Henry Gray's Anatomy of the Human Body*. http://en.wikipedia.org/wiki, 1858.

[59] H. Groemer. *Geometric Applications of Fourier Series and Spherical Harmonics,*. Cambridge University Press, 1996.

[60] X. Gu, Y.L. Wang, T.F. Chan, T.M. Thompson, and S.T. Yau. Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Transactions on Medical Imaging*, 23:1–10, 2004.

[61] J.F. Hair, R.L Tatham, R.E. Anderson, and W.C. Black. *Multivariate Data Analysis*. Prentice Hall, Inc., 1998.

[62] AY Hardan, NJ Minshew, and MS Keshavan. Corpus callosum size in autism. *Neurology*, 55:1033, 2000.

[63] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*. Springer, 2003.

[64] R. Higdon, N.L. Foster, R.A. Koeppe, C.S. DeCarli, W.J. Jagust, C.M. Clark, N.R. Barbas, S.E. Arnold, J.L. Heidebrink R.S. Turner, and S. Minoshima. A comparison of classification methods for differentiating fronto-temporal dementia from alzheimer's disease using FDG-PET imaging. *Stat Med*, 23:315–326, 2004.

[65] T.J. Hoffmann, M.K. Chung, K.D. Dalton, A.L. Alexander, G. Wahba, and R.J. Davidson. Subpixel curvature estimation of the corpus callosum via splines and its application to autism. In *10th Annual Meeting of the Organization for Human Brain Mapping*, 2004.

[66] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, London, 1985.

[67] A.J. Jerri. *The Gibbs phenomenon in Fourier analysis, splines and wavelet approximations*. Springer, 1998.

[68] S. Joshi, S. Pizer, PT Fletcher, P. Yushkevich, A. Thall, and JS Marron. Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. *IEEE Transactions on Medical Imaging*, 21:538–550, 2002.

[69] S.C. Joshi. *Large Deformation Diffeomorphisms and Gaussian Random Fields for Statistical Characterization of Brain Sub-Manifolds*. 1998.

[70] S.C. Joshi, U. Grenander, and M.I. Miller. The geometry and shape of brain sub-manifolds. *International Journal of Pattern Recognition and Artificial Intelligence: Special Issue on Processing of MR Images of the Human*, 11:1317–1343, 1997.

[71] S.C. Joshi, J. Wang, M.I. Miller, D.C. Van Essen, and U. Grenander. On the differential geometry of the cortical surface. *Vision Geometry IV, Vol. 2573, Proceedings of the SPIE's 1995 International Symposium on Optical Science, Engineering and Instrumentation*, pages 304–311, 1995.

[72] A. Kelemen, G. Szekely, and G. Gerig. Elastic model-based segmentation of 3-d neuroradiological data sets. *IEEE Transactions on Medical Imaging*, 18:828–839, 1999.

[73] D.N. Kennedy, K.M. O'Craven, B.S. Ticho, A.M. Goldstein, N. Makris, and J.W. Henson. Structural and functional brain asymmetries in human situs inversus totalis. *Neurology*, 53:1260–1265, 1999.

[74] S. Kovačič and R. Bajcsy. Multiscale/multiresolution representations. *Brain Warping*, pages 45–65, 1999.

[75] S. Kwapien and W.A. Woyczynski. *Random Series and Stochastic Integrals: Single and Multiple*. Probability and Its Applications. Birkhauser, 1992.

[76] M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York, 1982.

[77] J. P. Lerch and A.C. Evans. Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage*, 24:163–173, 2005.

[78] M.E. Leventon, W.E.L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, 2000.

[79] T. Lindeberg. *Scale-Space Theory in Computer Vision.* Kluwer Academic Publisher, 1994.

[80] E. Luders, Thompson, P.M., K.L. Narr, A.W. Toga, L. Jancke, and C. Gaser. A curvature-based approach to estimate local gyrification on the cortical surface. *NeuroImage*, 29:1224–1230, 2006.

[81] J.D. MacDonald, N. Kabani, D. Avis, and A.C. Evans. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from mri. *NeuroImage*, 12:340–356, 2000.

[82] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41:3397–3415, 1993.

[83] F. Manes, J. Piven, D. Vrancic, V. Nanclares, C. Plebst, and S.E. Starkstein. An MRI study of the corpus callosum and cerebellum in mentally retarded autistic individuals. *Journal of Neuropsychiatry and Clinical Neurosciences*, 11:470, 1999.

[84] M. Mather, T. Canli, T. English, S. Whitfield, P. Wais, K. Ochsner, J.D.E. Gabrieli, and L.L. Carstensen. Amygdala responses to emotionally valenced stimuli in older and younger adults. *Psychological Science*, 15:259–263, 2004.

[85] M.I. Miller, A. Banerjee, G.E. Christensen, S.C. Joshi, N. Khaneja, U. Grenander, and L. Matejic. Statistical methods in computational anatomy. *Statistical Methods in Medical Research*, 6:267–299, 1997.

[86] M.I. Miller, A.B. Massie, J.T. Ratnanather, K.N. Botteron, and J.G. Csernansky. Bayesian construction of geometrically based cortical thickness metrics. *NeuroImage*, 12:676–687, 2000.

[87] H.-G. Muller. Functional modeling and classification of longitudinal data. *Scandinavian Journal of Statistics*, 32:223–240, 2005.

[88] D. Nain, M. Styner, M. Niethammer, J.J. Levitt, M.E. Shenton, G. Gerig, A. Bobick, and A. Tannenbaum. Statistical shape analysis of brain structures using spherical wavelets. In *IEEE Symposium on Biomedical Imaging ISBI*, 2007.

[89] T. Nichols and S. Hayasaka. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Methods Med. Res.*, 12:419–446, 2003.

[90] T.E. Nichols and A.P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, 15:1–25, 2002.

[91] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[92] J.C. Pinehiro and D.M. Bates. *Mixed Effects Models in S and S-Plus*. Springer, 3rd. edition, 2002.

[93] J. Piven, S. Arndt, J. Bailey, and N. Andreasen. Regional brain enlargement in autism: a magnetic resonance imaging study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 35:530–536, 1996.

[94] J. Piven, J. Bailey, B.J. Ranson, and S. Arndt. An MRI study of the corpus callosum in autism. *American Journal of Psychiatry*, 154:1051, 1997.

[95] S.M. Pizer, D.S. Fritsch, P.A. Yushkevich, V.E. Johnson, and E.L. Chaney. Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Transactions on Medical Imaging*, 18:851–865, 1999.

[96] J-B Poline and B.M. Mazoyer. Analysis of individual brain activation maps using hierarchical description and multiscale detection. *IEEE Transactions on Medical Imaging*, 13:702–710, 1994.

[97] J-B Poline, K.J. Worsley, Frackowiak R.S.J. Holmes, A.P., and K.J. Friston. Estimating smoothness in statistical parametric maps: Variability of p values. *Journal of Computer Assisted Tomography*, 19:788–796, 1995.

[98] A. Qiu, D. Bitouk, and M.I. Miller. Smooth functional and structural maps on the neocortex via orthonormal bases of the laplace-beltrami operator. *IEEE Transactions on Medical Imaging*, 25:1296–1396, 2006.

[99] M. Quicken, C Brechbuhler, J. Hug, H. Blattmann, and G. Szekely. Parameterization of closed surfaces for parametric surface description. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 354–360, 2000.

[100] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer-Verlag, 1997.

[101] S.O. Rice. Mathematical analysis of random noise. *Bell System Tech. J*, 23:282–332, 1944.

[102] S.M. Robbins. Anatomical standardization of the human brain in euclidean 3-space and on the cortical 2-manifold. Technical Report PhD thesis, School of Computer Science, McGill University, Montreal, Quebec, Canada, 2003.

[103] C.P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Verlag, 2004.

[104] W. Rudin. *Functional Analysis*. McGraw-Hill, 1991.

[105] V. Schmidt and E. Spodarev. Joint estimators for the specific intrinsic volumes of stationary random sets. *Stochastic Processes and their Applications*, 115:959–981, 2005.

[106] L. Shen and M.K. Chung. Large-scale modeling of parametric surfaces using spherical harmonics. In *Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, 2006.

[107] L. Shen, J. Ford, F. Makedon, and A. Saykin. surface-based approach for classification of 3d neuroanatomical structures. *Intelligent Data Analysis*, 8:519–542, 2004.

[108] D.O. Siegmund and K.J. Worsley. Testing for a signal with unknown location and scale in a stationary gaussian random field. *Annals of Statistics*, 23:608–639, 1996.

[109] C.F. Stevens. *Introduction to Linear Algebra, 3rd edition*. Wellesley-Cambridge Press, 2003.

[110] M. Styner, G. Gerig, S. Joshi, and S. Pizer. Automatic and robust computation of 3d medial models incorporating object variability. *International Journal of Computer Vision*, 55:107–122, 2003.

[111] M. Styner, I. Oguz, S. Xu, C. Brechbuhler, D. Pantazis, J. Levitt, M. Shenton, and G. Gerig. Framework for the statistical shape analysis of brain structures using spharm-pdm. In *Insight Journal, Special Edition on the Open Science Workshop at MICCAI*, 2006.

[112] J. Talairach and P. Tournoux. Co-planar stereotactic atlas of the human brain: 3-dimensional proportional system. an approach to cerebral imaging. *Thieme, Stuttgart*, 1988.

[113] J.E. Taylor and K.J. Worsley. Detecting sparse signals in random fields, with an application to brain mapping. *Journal of the American Statistical Association*, 102:913–928, 2007.

[114] J.E. Taylor and K.J. Worsley. Random fields of multivariate test statistics, with applications to shape analysis. *Annals of Statistics*, 36:1–27, 2008.

[115] C. Thomaz, J.P. Boardman, S.J. Counsell, D.L.G. Hill, J.V. Hajnal, A.D. Edwards, M.A. Rutherford, D. Gillies, and D. Rueckert. A whole brain morphometric analysis of changes associated with preterm birth. In *SPIE Medical Imaging 2006: Image Processing*, volume 6144, pages 1903–1910, 2006.

[116] P. M. Thompson, D. MacDonald, M. S. Mega, C. J. Holmes, A. C. Evans, and A. W Toga. Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. *Journal of Computer Assisted Tomography*, 21:567–581, 1997.

[117] P.M. Thompson, J.N. Giedd, R.P. Woods, D. MacDonald, A.C. Evans, and A.W Toga. Growth patterns in the developing human brain detected using continuum-mechanical tensor mapping. *Nature*, 404:190–193, 2000.

[118] P.M. Thompson and A.W. Toga. A surface-based technique for warping 3-dimensional images of the brain. *IEEE Transactions on Medical Imaging*, 15, 1996.

[119] N.H. Timm and T.A. Mieczkowski. *Univariate and multivariate general linear models : theory and applications using SAS software*. SAS Publishing, 1997.

[120] A.W. Toga and P.M. Thompson. Mapping brain asymmetry. *Nature Reviews Neuroscience*, 4:37–48, 2003.

[121] C.N. Vidal, T.J. DeVito, K.M. Hayashi, D.J. Drost, P.C. Williamson, B. Craven-Thuss, D. Herman, Y. Sui, A.W. Toga, R. Nicolson, and P.M. Thompson. Detection and visualization of corpus callosum deficits in autistic children using novel anatomical mapping algorithms. In *Proc. International Society for Magnetic Resonance in Medicine*, 2003.

[122] G. Wahba. *Spline models for observational data*. SIAM, 1990.

[123] H. Wilbraham. On a certain periodic function. *Cambridge and Dublin Math. Journal*, 3:198–201, 1848.

[124] S.F. Witelson. The brain connection: the corpus callosum is larger in left-handers. *Science*, 229:665, 1985.

[125] S.F. Witelson. Hand and sex differences in the isthmus and genu of the human corpus callosum: a postmortem morphological study. *Brain*, 112:799, 1989.

[126] A. Witkin. Scale-space filtering. In *Int. Joint Conference on Artificial Intelligence*, pages 1019–1021, 1983.

[127] K.J. Worlsey, J-B. Poline, A.C. Vandal, and K.J. Friston. Test for distributed, non-focal brain activations. *NeuroImage*, 2:173–181, 1995.

[128] K.J. Worsley. Local maxima and the expected euler characteristic of excursion sets of $\chi 2$, f and t fields. *Advances in Applied Probability.*, 26:13–42, 1994.

[129] K.J. Worsley. Detecting activation in fmri data. *Statistical Methods in Medical Research.*, 12:401–418, 2003.

[130] KJ Worsley, J. Cao, T. Paus, M. Petrides, and AC Evans. Applications of random field theory to functional connectivity. *Hum Brain Mapp*, 6:364–7, 1998.

[131] K.J. Worsley, AC Evans, S. Marrett, and P. Neelin. A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12:900–900, 1992.

[132] K.J. Worsley, S. Marrett, P. Neelin, and A.C. Evans. Searching scale space for activation in pet images. *Human Brain Mapping*, 4:74–90, 1996.

[133] K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston, and A.C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4:58–73, 1996.

[134] K.J. Worsley, J.E. Taylor, F. Tomaiuolo, and J. Lerch. Unified univariate and multivariate random field theory. *NeuroImage*, 23:S189–195, 2004.

[135] A.M. Yaglom. *Correlation Theory of Stationary and Related Random Functions Vol. I: Basic Results.* Springer-Verlag, 1987.

[136] P. Yu, P.E. Grant, Y. Qi, X. Han, F. Segonne, R. Pienaar, E. Busa, J. Pacheco, N. Makris, R.L. Buckner, et al. Cortical Surface Shape Analysis Based on Spherical Wavelets. *IEEE Transactions on Medical Imaging*, 26:582, 2007.