

DEPARTMENT OF STATISTICS  
University of Wisconsin  
1210 West Dayton St.  
Madison, WI 53706

TECHNICAL REPORT NO. 1090  
July 29, 2004

Heat Kernel Smoothing and  
its Application to Cortical Manifolds

Moo K. Chung

Department of Statistics  
Department of Biostatistics and Medical Informatics  
Keck Laboratory for Functional Brain Imaging and Behavior  
University of Wisconsin-Madison  
Madison, WI 53706

# Heat Kernel Smoothing and its Application to Cortical Manifolds

Moo K. Chung  
Department of Statistics  
Department of Biostatistics and Medical Informatics  
Keck Laboratory for Functional Brain Imaging and Behavior  
University of Wisconsin-Madison  
Madison, WI 53706  
mchung@stat.wisc.edu

July 29, 2004

## **Abstract**

In brain imaging analysis, there is a need for analyzing data collected on the cortical surface of the human brain. Gaussian kernel smoothing has been widely used in this area in conjunction with random field theory for analyzing data residing in Euclidean spaces. The Gaussian kernel is isotropic in Euclidian space so it assigns the same weights to observations equal distance apart. However, when we smooth data residing on a curved surface, it fails to be isotropic. On the curved surface, a straight line between two points is not the shortest distance so one may assign smaller weights to closer observations. For this reason smoothing data residing on manifolds requires constructing a kernel that is isotropic along the geodesic curves. With this motivation in mind, we construct the kernel of a heat equation on manifolds that should be isotropic in the local conformal coordinates and develop a framework for heat kernel smoothing and statistical inference is performed on manifolds. As an illustration, we apply our approach in comparing the cortical thickness of autistic children to that of normal children.

# 1 Introduction

The cerebral cortex has the topology of a 2D highly convoluted sheet. Most of the anatomical features that distinguish these cortical regions can only be measured relative to the cortical surface. It is likely that different clinical population will show different brain surface shape differences. By measuring the cortical thickness difference among groups, brain shape differences can be quantified locally.

The cortical surface is usually represented as a triangular mesh with average inter nodal distance of 1-3mm (Figure 1). The most widely used method for triangulating the cortical surface is the deformable surfaces method (MacDonald *et al.*, 2000). It can generate cortical triangular meshes that has the topology of a sphere consisting of 40,962 vertices and 81,920 triangles with the average internodal distance of 3 mm. Once we have a triangular mesh as the realization of the cortical surface, we can compute the distance between two cortical boundaries (Figure 2). It is natural to assume the cortical surfaces to be a smooth 2-dimensional Riemannian manifold (Dale and Fischl, 1999; Joshi *et al.*, 1995).

To increase the signal to noise ratio and smoothness, diffusion smoothing, which generalizes Gaussian kernel smoothing to an arbitrary manifolds, has been developed and have been used in brain imaging (Andrade *et al.* 2001; Cachia *et al.*, 2003; Chung *et al.*, 2003). Due to huge noise on the cortical thickness measurements, it is necessary to smooth it along the cortical surface (Figure 3). The smoothing is also necessary to grantee the smoothness in the random fields theory.

Consider following stochastic model for thickness on the cortex  $\partial\Omega$ :

$$Y(p) = \theta(p) + \epsilon(p), p \in \partial\Omega \quad (1)$$

where  $Y$  is thickness measurement,  $\theta$  true unknown thickness and  $\epsilon$  is a zero mean Gaussian random field. Then the solution to a diffusion equation is used as an estimate of  $\theta$ . The drawback of this method is the complexity of the setting up a finite element method (FEM) for solving the diffusion equation numerically. To overcome this shortcoming associated with solving diffusion equation on manifolds, we have developed a much simpler method based on the heat kernel convolution which generalizes Gaussian kernel smoothing in Euclidean space to arbitrary Riemannian manifolds.

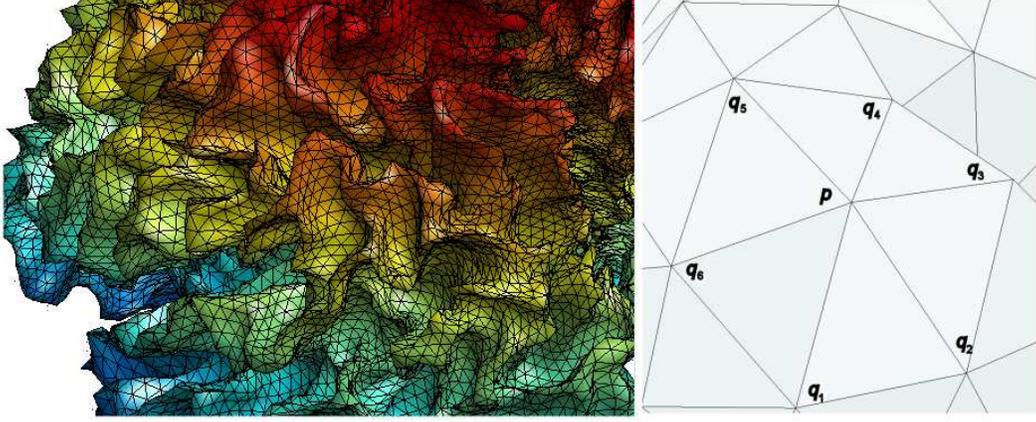


Figure 1: Right: Typical triangular surface representation of the brain cortex. Left: Typical triangular surface with  $m = 6$  neighboring vertices around  $p = q_0$

## 2 Heat Kernel Smoothing

We define *heat kernel smoothing estimator* of data  $\theta$  to be the convolution

$$\widehat{\theta}(p) = K_\sigma * Y(p) = \int_{\partial\Omega} K_\sigma(p, q) Y(q) d\mu(q) \quad (2)$$

where  $\mu(q)$  is a surface measure mostly the Lebesgue measure if not stated otherwise and heat kernel  $K_\sigma$  is given in terms of the eigenvalues of the Laplace-Beltrami operator. For an overview of heat kernel, one may refer Rosenberg (1997) and Berline *et al.* (1991). The Laplace-Beltrami operator  $\Delta$  corresponding to the surface parameterization  $p = X(u^1, u^2) \in \partial\Omega$  can be written as

$$\Delta = \frac{1}{\det g^{1/2}} \sum_{i,j=1}^2 \frac{\partial}{\partial u^i} \left( \det g^{1/2} g^{ij} \frac{\partial}{\partial u^j} \right)$$

where  $g = (g_{ij})$  is the Riemannian metric tensor given by inner product  $g_{ij} = \langle \frac{\partial X}{\partial u^i}, \frac{\partial X}{\partial u^j} \rangle$ . Let  $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \dots$  be ordered eigenvalues and  $\psi_0, \psi_1, \psi_2, \dots$  be the corresponding eigenfunction of the Laplace-Beltrami operator given by solving  $\Delta\psi_j = \lambda_j\psi_j$ . Note that  $\psi_j$  form an orthonormal basis of  $L^2$  space on manifolds  $\partial\Omega$ , i.e.  $L^2(\partial\Omega)$ . Assuming the existence of heat kernel, it has a spec-

tral representation

$$K_\sigma(p, q) = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} \psi_j(p) \psi_j(q). \quad (3)$$

Under some regularity condition,  $K_\sigma$  is a probability distribution on  $\partial\Omega$  which is a generalized version of Gaussian density on manifolds so  $\int_{\partial\Omega} K_\sigma(p, q) d\mu(q) = 1$  for all  $p \in \partial\Omega$  and  $\sigma \in \mathbb{R}^+$ . Another property of heat kernel is  $K_\sigma(p, q) = K_\sigma(q, p)$ . Convolution (2) can be viewed as the unique solution of a partial differential equation (PDE) and the minimizer of the weighted least-squares errors.

**Theorem 1**  $K_\sigma * Y$  is the unique solution of the following initial value problem at time  $t = \sigma^2/2$ :

$$\frac{\partial f}{\partial t} = \Delta f, \quad f(p, 0) = Y(p) \quad (4)$$

where  $\Delta$  is the Laplace-Beltrami operator.

This is a well known result in differential geometry (Rosenberg, 1997). In Chung *et al.* (2003), smoothing is performed by solving the above heat equation (4) via the finite element method.

**Theorem 2**  $K_\sigma * Y$  is the minimizer

$$K_\sigma * Y(p) = \arg \min_{\lambda} \int_{\partial\Omega} K_\sigma(p, q) [Y(q) - \theta]^2 d\mu(q).$$

Similar result is given for Gaussian kernel smoothing in the Euclidean space (Fan and Gijbels, 1996). It can be proved easily noting that the right hand side is quadratic in  $\theta$ :

$$K_\sigma * Y^2(p) - 2\theta K_\sigma * Y(p) + \theta^2$$

which is minimized at its extreme value. So the heat kernel smoothing can be viewed as the 0-th order weighted polynomial regression on manifolds.

**Theorem 3** Suppose the covariance function of  $Y$  in (1) is decreasing isotropic function, i.e.  $R_Y(p, q) = \rho(d(p, q))$  where  $d(p, q)$  is the geodesic distance between  $p$  and  $q$ . Then

$$\mathbf{Var}[K_\sigma * Y(p)] \leq \mathbf{Var}Y(p) \text{ for each } p \in \partial\Omega.$$

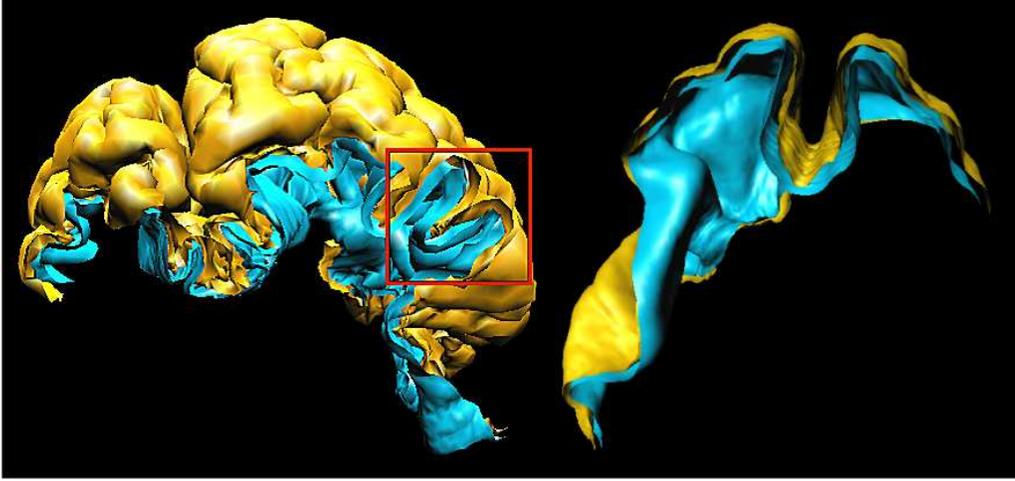


Figure 2: Anatomy of brain cortex. Left: Part of the cortical surface showing both outer (yellow) and inner surface (blue) that bound gray matter. Right: enlargement of the boxed region. The cortical thickness measures the distance between outer and inner surfaces.

The geodesic distance is defined in the following way. Consider curve segment  $C \subset \partial\Omega$  connecting  $p$  and  $q$  and parameterized by  $\gamma_c(t)$  with  $\gamma_c(0) = p$  and  $\gamma_c(1) = q$ . In Cartesian coordinates,  $\gamma_c(t) = (\gamma_c^1(t), \dots, \gamma_c^n(t)) \in \mathbb{R}^n$ . The length of  $C$  is given by

$$\int_0^1 \left\langle \frac{d\gamma_c}{dt}, \frac{d\gamma_c}{dt} \right\rangle^{1/2} dt = \int_0^1 \left[ \sum_{i,j} g_{ij} \frac{d\gamma_c^i}{dt} \frac{d\gamma_c^j}{dt} \right]^{1/2} dt$$

where the inner product  $\langle \cdot, \cdot \rangle$  is with respect to the tangent space of the manifold. Then the geodesic curve connecting  $p$  and  $q$  is defined as the minimizer

$$d(p, q) = \min_C \int_0^1 \left\langle \frac{d\gamma_c}{dt}, \frac{d\gamma_c}{dt} \right\rangle^{1/2} dt.$$

It is usually given as a solution the an Euler equation and computational technique is available for polygonal surfaces (Wolfson and Schwartz, 1989). Note that  $R_Y(p', q') = \rho(d(p', q')) \leq \rho(0) = \rho(d(p', p')) = R_Y(p', p') = \mathbf{Var}Y(p')$ . Isotropic covariance function implies stationary uniform variance field. The co-

variance function  $R$  of  $K_\sigma * Y(p)$  is given by

$$\begin{aligned}
R(p, q) &= \mathbb{E} \left[ \int_{\partial\Omega} K_\sigma(p, p') Y(p') d\mu(p') \right]^2 \\
&= \int_{\partial\Omega} \int_{\partial\Omega} K_\sigma(p, p') K_\sigma(q, q') R_Y(p', q') d\mu(p') d\mu(q') \\
&\leq \int_{\partial\Omega} \int_{\partial\Omega} K_\sigma(p, p') K_\sigma(q, q') \rho(0) d\mu(p') d\mu(q') \\
&= \rho(0)
\end{aligned}$$

from the fact that  $K_\sigma$  is a probability distribution. Now letting  $p = q$ , we have  $\mathbf{Var}[K_\sigma * Y(p)] = R(p, p) \leq \rho(0) \leq \mathbf{Var}Y(p)$  proving the theorem. Hence heat kernel smoothing will reduce the variability of cortical thickness measurements.

**Theorem 4** *Heat kernel smoothing with large bandwidth can be decomposed into multiple kernel smoothing with smaller bandwidth via*

$$K_\sigma^{(k)} * f = \underbrace{K_\sigma * \dots * K_\sigma}_{k \text{ times}} * f = K_{\sqrt{k}\sigma} * f.$$

From Theorem 1,  $K_\sigma * (K_\sigma * Y)$  can be taken as the diffusion of signal  $K_\sigma * Y$  after time  $\sigma^2/2$  so that  $K_\sigma * (K_\sigma * Y)$  is the diffusion of signal  $Y$  after time  $\sigma^2$ , i.e.

$$K_\sigma * K_\sigma * Y = K_{\sqrt{2}\sigma} * Y.$$

Arguing inductively, we prove the general statement.

In order to implement heat kernel smoothing numerically, we use asymptotic representation called the *parametrix expansion* (Rosenberg, 1997):

$$K_\sigma(p, q) = \frac{1}{(2\pi\sigma)^{1/2}} \exp \left[ -\frac{d^2(p, q)}{2\sigma^2} \right] [u_0(p, q) + O(\sigma^2)] \quad (5)$$

where  $d(p, q)$  is the geodesic distance between  $x$  and  $y$ . The first term  $u_0(p, q) \approx \det g^{-1/2}(q)$  for  $p$  close to  $q$  so that  $u_0(p, q) \rightarrow 1$  as  $p \rightarrow q$ . When the manifolds is flat,  $g_{ij} = \delta_{ij}$  and  $d(p, q) = \|p - q\|$ , the Euclidean distance between  $p$  and  $q$  so the heat kernel  $K_\sigma$  becomes Gaussian kernel

$$G_\sigma(p, q) = \frac{1}{(2\pi\sigma)^{1/2}} \exp \left[ -\frac{\|p - q\|^2}{2\sigma^2} \right].$$

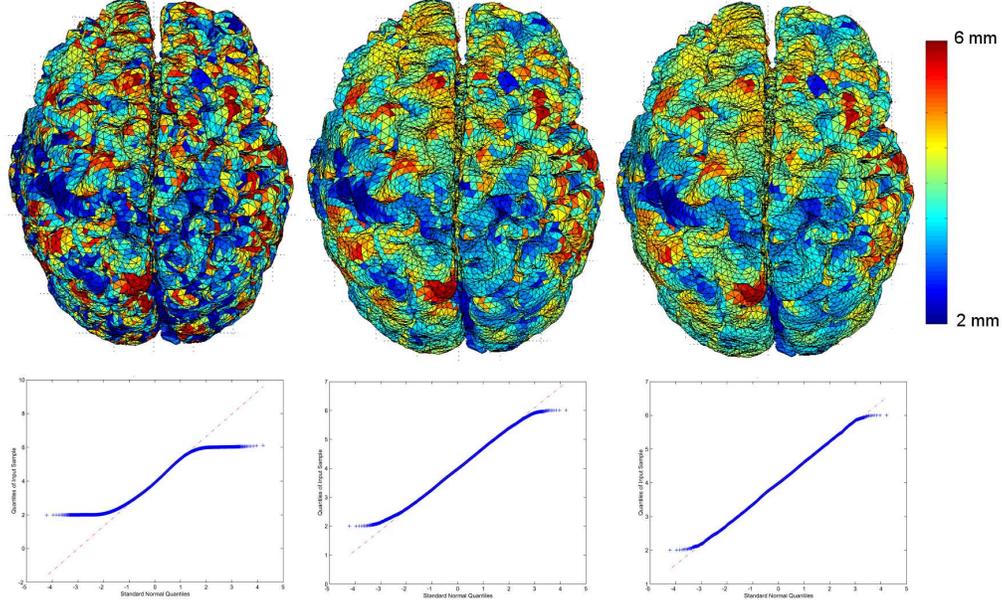


Figure 3: Left: original cortical thickness measurement and corresponding QQ-plot. Middle: after 50 iterations with smoothing parameter  $\sigma = 0.5\text{mm}$ ., Right: after 100 iterations. The total effective smoothing amount after 100 iterations is 5 mm. QQ-plots show increased Gaussianity.

Assuming sufficiently small  $\sigma$  and close  $p$  and  $q$ , we have

$$K_\sigma(p, q) \approx \frac{1}{(2\pi\sigma)^{1/2}} \exp\left[-\frac{d^2(p, q)}{2\sigma^2}\right].$$

But since the above kernel may not integrate to 1, we normalize kernel in a small geodesic ball  $B_p = \{q \in \partial\Omega : d(p, q) \leq r\} \subset \partial\Omega$ :

$$\tilde{K}_\sigma(p, q) = \frac{\exp\left[-\frac{d^2(p, q)}{2\sigma^2}\right] \mathbf{1}_{B_p}(q)}{\int_{B_p} \exp\left[-\frac{d^2(p, q)}{2\sigma^2}\right] d\mu(q)} \quad (6)$$

indicator function  $\mathbf{1}_{B_p}$  is defined as  $\mathbf{1}_{B_p}(q) = 1$  if  $q \in B_p$  and  $\mathbf{1}_{B_p}(q) = 0$  otherwise. Then  $\tilde{K}_\sigma(p, q)$  is a probability distribution on  $\partial\Omega$ . Since  $d\mu(q) = \det g^{1/2}(q) du^1 du^2$ , heat kernel smoothing with kernel (6) in parameter space is

given by

$$\tilde{K}_\sigma * Y(p) = \frac{\int_{B_p} \exp\left[-\frac{d^2(p,q)}{2\sigma^2}\right] \det g^{1/2}(q) Y(q) du^1 du^2}{\int_{B_p} \exp\left[-\frac{d^2(p,q)}{2\sigma^2}\right] \det g^{1/2}(q) du^1 du^2}.$$

**Theorem 5** *When the radius  $r$  of geodesic ball  $B_p$  is sufficiently large, heat kernel smoothing with infinite bandwidth gives average signal, i.e.*

$$\lim_{\sigma \rightarrow \infty} \tilde{K}_\sigma * Y = \frac{\int_{\partial\Omega} Y(q) d\mu(q)}{\mu(\partial\Omega)}.$$

From definition (6),

$$\lim_{\sigma \rightarrow \infty} \tilde{K}_\sigma = \frac{\mathbf{1}_{B_p}(q)}{\int_{B_p} d\mu(q)} = \frac{\mathbf{1}_{B_p}(q)}{\mu(B_p)}.$$

Hence

$$\lim_{\sigma \rightarrow \infty} \tilde{K}_\sigma * Y(p) = \frac{\int_{B_p} Y(q) d\mu(q)}{\mu(B_p)}$$

which is the average signal over  $B_p$ . Now letting the radius of the geodesic ball to be large enough that  $B_p = \partial\Omega$ , we prove the theorem. A similar result for heat kernel (3) is given in Rosenberg (1997).

**Theorem 6** *When  $\partial\Omega$  is a flat Euclidean space,*

$$G_{\sqrt{k}\sigma}(p) \leq \tilde{K}_\sigma^{(k)}(p) \leq \frac{1}{\alpha^k(B_p)} G_{\sqrt{k}\sigma}(p)$$

where  $\alpha(B_p) = \int_{B_p} G_\sigma(p, q) d\mu(q)$ .

Unlike heat kernel, truncated kernel does not have the nice property of Theorem 4, i.e.

$$\tilde{K}_\sigma * \tilde{K}_\sigma \neq \tilde{K}_{\sqrt{2}\sigma}.$$

In order to implement iterated kernel smoothing, we need a similar result. When the manifold is flat, truncated kernel (6) becomes

$$\tilde{K}_\sigma(p, q) = \frac{G_\sigma(p, q) \mathbf{1}_{B_p}(q)}{\int_{B_p} G_\sigma(p, q) d\mu(q)}.$$

Hence

$$\begin{aligned}
G_\sigma(p, q) &= G_\sigma(p, q)\mathbf{1}_{B_p}(q) + G_\sigma(p, q)\mathbf{1}_{\partial\Omega \setminus B_p}(q) \\
&= \alpha(B_p)\tilde{K}_\sigma(p, q) + G_\sigma(p, q)\mathbf{1}_{\partial\Omega \setminus B_p}(q) \\
&\geq \alpha(B_p)\tilde{K}_\sigma(p, q).
\end{aligned}$$

Applying convolution to the above equation, we have

$$\begin{aligned}
G_\sigma * G_\sigma(q) &\geq \alpha G_\sigma * \tilde{K}_\sigma(q) \\
&\geq \alpha^2 \tilde{K}_\sigma * \tilde{K}_\sigma(q).
\end{aligned}$$

Trivially  $G_\sigma(q) \leq \tilde{K}_\sigma(q)$  for  $q \in B_p$ . Hence

$$G_\sigma * G_\sigma(q) \leq \tilde{K}_\sigma * \tilde{K}_\sigma(q) \leq \frac{1}{\alpha^2(B_p)} G_\sigma * G_\sigma(q), q \in B_p$$

and the general result follows so that

$$G_\sigma^{(k)}(q) \leq \tilde{K}_\sigma^{(k)}(q) \leq \frac{1}{\alpha^k(B_p)} G_\sigma^{(k)}(q), q \in B_p.$$

Now applying the result of Theorem 4, we prove the statement.

Note that  $\alpha(B_p) \rightarrow 1$  as  $\sigma \rightarrow 0$ . So for the proper choice of sufficiently small  $\sigma$  and  $B_p$ , we can make  $\alpha(B_p)$  as small as we want. Suppose  $\alpha = 0.9999$ . The number of iterations will be fixed to be less than 100 or 200 in numerical implementation. That gives  $\alpha^{100} = 0.9900$  and  $\alpha^{200} = 0.9802$ . For  $\alpha = 0.999$ ,  $\alpha^{100} = 0.9048$  and  $\alpha^{200} = 0.8186$ . Hence, decreasing the size of bandwidth and increasing the number of iterations would perform better. This theorem shows that the iterated truncated kernel smoothing can approximate the integral version of kernel smoothing within one percent error if one desires it in practical applications. Theorem 4 should be asymptotically true for general manifolds when  $B_p$  and  $\sigma$  are sufficiently small.

In the case of triangular mesh, cortical thickness is measured at discrete vertices so it is natural to take a discrete measure  $\mu$  in defining convolution. Let  $q_1, \dots, q_m$  be neighboring vertices of  $p = q_0$  and  $N_p = \{q_0, q_1, \dots, q_m\}$  be the set of nearest neighboring points of  $p$  plus the point itself (Figure 1). Then the geodesic distance between adjacent vertices is the straight line, i.e.  $d(p, q_i) = \|p - q_i\|$ , the Euclidean distance. Then we define the normalized truncated kernel for polygonal surface to be

$$\tilde{W}_\sigma(p, q_i) = \frac{\exp\left[-\frac{d^2(p, q_i)}{2\sigma^2}\right]}{\sum_{j=0}^m \exp\left[-\frac{d^2(p, q_j)}{2\sigma^2}\right]}$$

and discrete convolution

$$\widetilde{W}_\sigma * Y(p) = \sum_{i=0}^m \widetilde{W}_\sigma(p, q_i) Y(q_i).$$

Note  $\widetilde{W}_\sigma$  is a discrete probability distribution so that  $\sum_{i=0}^m \widetilde{W}_\sigma(p, q_i) = 1$ . This is the generalization of *Nadaraya-Watson estimator* (Fan and Gijbels, 1996) for discrete measurements in manifolds. Following the proof of Theorem 4, we see that the same results should hold for the discrete version with the discrete measure  $\mu$  and geodesic ball  $B_p$  replaced by  $N_p$ . For the proof to hold, we require

$$\alpha(N_p) = \sum_{i=0}^m G_\sigma(p, q_i) \leq 1.$$

In related note, Lafferty and Lebanon (2004) estimated the Laplace-Beltrami operator using similar exponential weights. Let polygonal surface  $S$  have  $n$  vertices  $q_1, \dots, q_n$ . Then we have the following algorithm for the heat kernel smoothing

**Algorithm 1**

*For  $i = 1$  to  $n$  do*

*Find a set of neighboring vertices  $N(q_i)$  of  $q_i$ .*

*Compute the weighted average and store  $Z(q_i) \leftarrow W_\sigma * Y(q_i)$ .*

*End.*

*Update  $Y \leftarrow Z$ .*

*Repeat this procedures  $k$ -times.*

### 3 Statistical Inference on Manifolds

We let the first group to be autistic and the second group to be normal control. There are  $n_i$  subjects in  $i$ -th group. Following stochastic model (1) for  $i$ -th group, we have the following model on cortical thickness  $Y_{i_j}$  for  $i$ -th group and  $j$ -th subject:

$$Y_{i_j}(p) = \theta_i(p) + \epsilon_{i_j}(p)$$

where  $\epsilon_{i_j}$  is independent zero mean Gaussian random fields. Then we are interested in testing if the thickness for two groups are identical, i.e.

$$H_0 : \theta_1(p) = \theta_2(p) \text{ for all } p \in \partial\Omega$$

v.s.

$$H_1 : \theta_1(p) > \theta_2(p) \text{ for some } p \in \partial\Omega.$$

The above hull hypothesis is the intersection of collection of hypothesis

$$H_0 = \bigcap_{p \in \partial\Omega} H_0(p)$$

where  $H_0(p) : \theta_1(p) = \theta_2(p)$ . Assuming the variability of two groups are same, which will be demonstrated for our data in the result section, the test statistic to use is the two sample  $t$ -statistic with equal variance given by

$$T(p) = \frac{\bar{\theta}_1 - \theta_1 - (\bar{\theta}_2 - \theta_2)}{S_p \sqrt{1/m + 1/n}}$$

where the pooled variance  $S_p^2 = ((n_1 - 1)S_1^2 + (n_2 - 1)S_2^2)/(n_1 + n_2 - 2)$ . Under null hypothesis,  $T(p) \sim t_{n_1+n_2-2}$  the  $t$ -distribution with  $n_1 + n_2 - 2$  degrees of freedom at each fixed point  $p$ . The type I error for the multiple hypotheses testing would be

$$\begin{aligned} \alpha &= P(\text{reject at least one } H_0(p) | H_0 \text{ true}) \\ &= P\left(\bigcup_{p \in \partial\Omega} \{T(p) > h\}\right) \\ &= 1 - P\left(\bigcap_{p \in \partial\Omega} T(p) \leq h\right) \\ &= 1 - P(\sup_{p \in \partial\Omega} T(p) \leq h) \\ &= P(\sup_{p \in \partial\Omega} T(p) > h). \end{aligned}$$

So in order to construct  $\alpha$ -level test for multiple hypothesis testing, we need to know the distribution of the supremum of correlated  $t$ -field.  $P$ -value constructed this way is usually refereed as the *corrected P-value* to distinguish it from the  $P$ -value constructed from a single hypothesis (add reference). The distribution of  $\sup_{p \in \partial\Omega} T(p)$  is asymptotically given as

$$P(\sup_{p \in \partial\Omega} T(p) > h) \approx \sum_{d=0}^2 \phi_d(\partial\Omega) \rho_d(h)$$

where  $\mu_d$  are the  $d$ -dimensional Minkowski functionals of  $\partial\Omega$  and  $\rho_d$  are the  $d$ -dimensional Euler characteristic (EC) density of  $t$ -field (Worsley *et al.*, 1996).

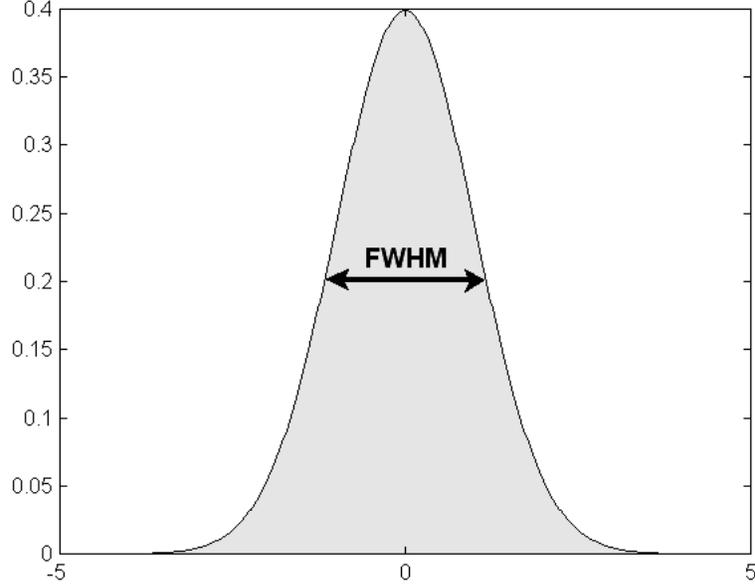


Figure 4: Full width at half maximum (FWHM) of Gaussian kernel.

The Minkowski functionals are  $\phi_0 = 2, \phi_1 = 0, \phi_2 = \mu(\partial\Omega)/2 = 49,616\text{mm}^2$ , the half area of the template cortex  $\partial\Omega$ . The EC density is given by

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

where dot notation indicates differentiation with respect to the first  $d$  components. For  $t$  random field with  $d$  degrees of freedom, the EC-densities are given by

$$\rho_0(h) = \int_h^\infty \frac{\Gamma(\frac{d+1}{2})}{(d\pi)^{1/2}\Gamma(\frac{d}{2})} \left(1 + \frac{x^2}{d}\right)^{-\frac{(d+1)}{2}} dx,$$

$$\rho_2(h) = \frac{\lambda}{(2\pi)^{3/2}} \frac{\Gamma(\frac{d+1}{2})}{(\frac{d}{2})^{1/2}\Gamma(\frac{d}{2})} \left(1 + \frac{h^2}{d}\right)^{-\frac{(d-1)}{2}}$$

where  $\lambda$  measures the smoothness of fields (Worsley *et al.*, 1994).

Similarly we can also perform a test based on the supremum of correlated  $F$ -fields as well. For removing the effect of age, we set up a general linear model (GLM) on cortical thickness  $Y_j$  for subject  $j$

$$Y_j(p) = \lambda_1(p) + \lambda_2(p) \cdot \text{age}_j + \beta(p) \cdot \text{group}_j + \epsilon$$

where dummy variable group is 1 for autism and 0 for normal control. Then we test the group difference

$$H_0 : \beta(p) = 0 \text{ for all } p \in \partial\Omega$$

v.s.

$$H_0 : \beta(p) \neq 0 \text{ for some } p \in \partial\Omega$$

Let us denote the sum of the squared errors (SSE) of the least squares estimation of parameters by

$$\text{SSE}_0(p) = \sum_{j=1}^{n_1+n_2} (Y_j(p) - \hat{\lambda}_1(p) - \hat{\lambda}_2(p) \cdot \text{age}_j)^2$$

and

$$\text{SSE}_1(p) = \sum_{j=1}^{n_1+n_2} (Y_j(p) - \hat{\lambda}_1(p) - \hat{\lambda}_2(p) \cdot \text{age}_j - \hat{\beta}(p) \cdot \text{group}_j)^2.$$

Then under  $H_0$ , the test statistic would follow  $F$  distribution at each fixed point  $p$ , i.e.

$$F(p) = \frac{\text{SSE}_0 - \text{SSE}_1}{\text{SSE}_0 / (n_1 + n_2 - 3)} \sim F_{1, n_1+n_2-3}$$

For  $F$  random field with  $\alpha$  and  $\beta$  degrees of freedom, the EC-densities are given by

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

where  $\lambda$  measures the smoothness of fields. If we assume  $\epsilon$  to be unit variance isotropic field, the smoothness isotropic random field can be defined as the covariance of derivative vector  $\partial\epsilon(x)/\partial x$ , given by  $\mathbf{Cov} \frac{\partial\epsilon(x)}{\partial x} = \lambda I$ . If  $\epsilon$  is the convolution of the Gaussian white noise with isotropic kernel  $K_\sigma$ , the covariance function of  $\epsilon$  is

$$R_\epsilon(x, y) = \int K_\sigma(x-z)K_\sigma(y-z) dz.$$

Since field  $\epsilon$  does not give unit variance, we normalize  $\epsilon$  by the square root of

$$\mathbf{Var} \epsilon = R_\epsilon(x, x) = \int K_\sigma^2(x - z) dz$$

The cross-covariance of the above zero mean unit variance field is then  $\lambda_{ij} = \mathbb{E}[\partial_{x_i} \epsilon \partial_{x_j} \epsilon] / \mathbb{E} \epsilon^2$ . Note that

$$\frac{\mathbb{E} \partial_{x_i} \epsilon(x) \partial_{y_j} \epsilon(y)}{\mathbb{E} \epsilon^2(x)} = \frac{\int \partial_{x_i} K_\sigma(x - z) \partial_{y_j} K_\sigma(y - z) dz}{\int K_\sigma(z) K_\sigma(z) dz}.$$

Now letting  $x = y$  and using identity

$$K_\sigma^2(x) = (2\sqrt{\pi}\sigma)^{-n} K_{\sigma/\sqrt{2}}(x),$$

we have

$$\lambda_{ij} = \frac{\int \frac{x_i^2}{\sigma^4} K_\sigma^2(x) dx}{\int K_\sigma^2(x) dx} \delta_{ij} = \frac{1}{2\sigma^2} \delta_{ij},$$

where  $\delta_{ij}$  is the Kronecker's delta. Hence  $\lambda = 1/(2\sigma^2)$ .

The amount of smoothing  $\sigma^2$  is usually expressed in terms of the full width at the half maximum (FWHM) of a smoothing kernel in brain imaging (Figure 4). Note that FWHM of kernel  $K_\sigma$  corresponds to  $2\sqrt{\ln 4}\sigma$ . Conversely for given FWHM, the corresponding kernel is  $K_{\text{FWHM}/(2\sqrt{\ln 4})}$ . So in terms of FWHM, the smoothness of field is given as  $\lambda = 4 \ln 2 / \text{FWHM}^2$ . The FWHM is usually predetermined to match the extend of the signal size and we set it to be 30 mm reflecting the width of sulci (Chung *et al.*, 2003).

## 4 Application

Gender and handedness affect brain anatomy (Luders, 2003) so all the 16 autistic and 12 control subjects used in this study were screened to be right-handed males except one subject who is ambidextrous. Sixteen autistic subjects were diagnosed with high functioning autism (HFA) via the Autism Diagnostic Interview - Revised (ADI-R) by a trained and certified psychologist (Chung *et al.*, 2003). 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  $17.1 \pm 2.8$  and the autistic subjects is  $16.1 \pm 4.5$  which is in compatible age range.

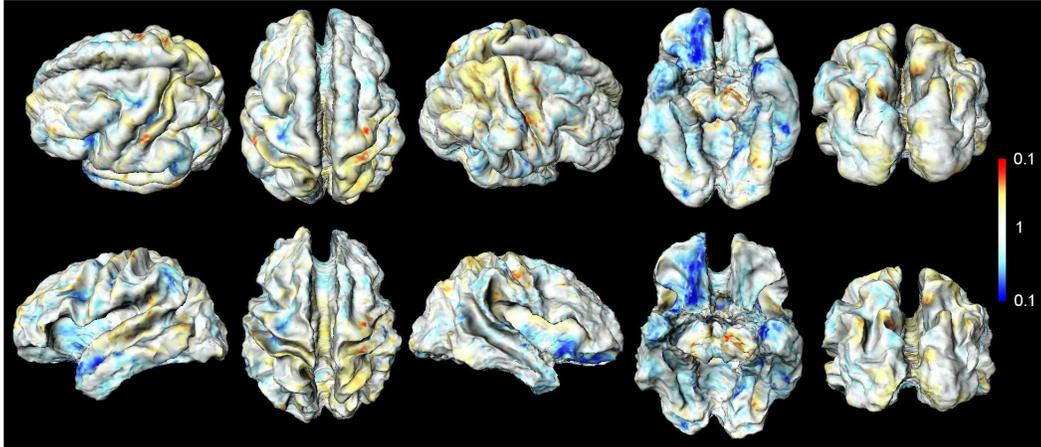


Figure 5: Corrected  $P$ -value map constructed from  $t$  random fields. Top is  $P$ -values projected onto the outer template surface and Bottom is  $P$ -values projected onto the inner template surface. Red is the regions of thicker gray matter while blue is thinner gray matter in the autistic subjects compared to the normal controls.

Afterwards, high resolution anatomical magnetic resonance images (MRI) were obtained using a 3-Tesla GE SIGNA (General Electric Medical Systems, Waukesha, WI) scanner with a quadrature head RF coil. The dimension of MRI is  $256 \times 256 \times 128$  with an approximate image resolution of  $1\text{mm}^3$ . For the detailed image acquisition parameters, see Chung *et al.* (2003) where the same data set is used to perform a different morphometric analysis. Afterwards, nonuniformity of image intensity has been corrected artifacts using nonparametric nonuniform intensity normalization method (N3), which eliminates the dependence of the field estimate on anatomy (Sled *et al.*, 1998). Then using the automatic image processing pipeline (Zijdenbos *et al.*, 1998), MRI were spatially normalized into a standardized template brain via a global affine transformation to align and remove the global brain volume difference (Collins *et al.*, 1994). Subsequently, an automatic tissue-segmentation algorithm based on a supervised artificial neural network classifier was used to classify each voxel as cerebrospinal fluid (CSF), gray matter, and white matter (Kollakian, 1996). Afterward, a triangular mesh for each cortical surface was generated by deforming a spherical mesh to fit the proper boundary in a segmented volume using a deformable surface algorithm (MacDonald *et al.*, 2000). Brain substructures such as the brain stem and the cerebellum were removed. Then an ellipsoidal mesh that already had the topology of a sphere

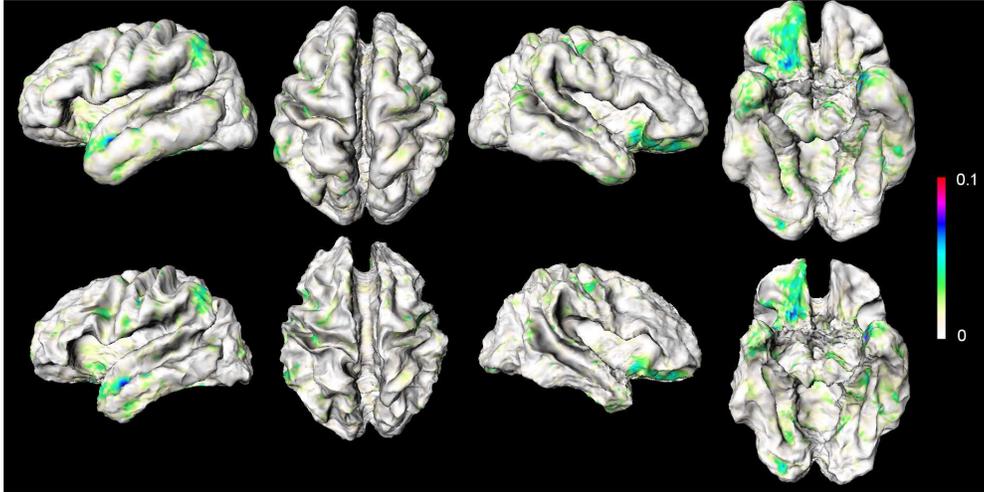


Figure 6: Corrected  $P$ -value map for  $F$  random fields removing the effects of age and the total gray matter volume.

was deformed to fit the shape of the cortex guaranteeing the same topology. The resulting triangular mesh consists of 40,962 vertices and 81,920 triangles with the average internodal distance of 3 mm. The triangular meshes are not constrained to lie on voxel boundaries. Instead the triangular meshes can cut through a voxel, which can be considered as minimizing the discretization error. Once we have a triangular mesh as the realization of the cortical surface, we compute the cortical thickness which measures the distance between the outer and inner surfaces that bound gray matter. Finally the thickness measurements are smoothed with a heat kernel of size 30 mm FWHM as described in a previous section and statistical analyses are performed and the final corrected  $P$ -value maps are computed (Figure 5 and 6). After removing the effect of age and the total gray matter volume, we found statistically thinner gray matter regions at the right temporal lobe and the left frontal lobe (corrected  $P$ -value  $< 0.1$ ).

## Acknowledgements

Kim Dalton and Richard Davidson of the Keck Laboratory for Functional Brain Imaging and Behavior, University of Wisconsin-Madison provided the autism data and Steve Robinson and Alan Evans of the Montreal Neurological Institute (MNI) assisted with the image segmentation. I would like to thank Keith

Worsley of the McGill University, Kam Tsui and Shijie Tang of the Department of Statistics, University of Wisconsin-Madison and Tulaya Limpiti of the Department of Electrical Engineering, University of Wisconsin-Madison for valuable discussions.

## References

1. Andrade, A., Kherif, F., Mangin, J., Worsley, K.J., Paradis, A., Simon, O., Dehaene, S., Le Bihan, D., and Poline, J-B. 2001. Detection of Fmri Activation Using Cortical Surface Mapping. *Human Brain Mapping*, **12**:79–93.
2. Berline, N., Getzler, E., and Vergne M. 1991. *Heat Kernels and Dirac Operators*, Springer-Verlag, New York.
3. Cachia, A., Mangin, J.-F., Riviere, D., Kherif, F., Boddaert, N., Andrade, A., Papadopoulos-Orfanos, D., Poline, J.-B., Bloch, I., Zilbovicius, M., Sonigo, P., Brunelle, F., Regis. 2003. A primal sketch of the cortex mean curvature: a morphogenesis based approach to study the variability of the folding patterns, **22**:754- 765.
4. Chung, M.K., Worsley, K.J., Robbins, S., Paus, P., Taylor, J., Giedd, J.N., Rapoport, J.L., Evans, A.C. 2003. Deformation-Based Surface Morphometry with an Application to Gray Matter Deformation, *NeuroImage*. **18**:198–213.
5. Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C. 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assisted Tomogr.* **18**:192-205.
6. Dale, A.M., and Fischl, B. 1999. Cortical surface-based analysis I. segmentation and surface reconstruction. *NeuroImage*, 9:179.194.
7. Fan, J., and Gijbels, I. 1996. *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC.
8. Joshi, S.C., Wang, J. and Miller, M.I., Van Essen, D.C., Grenander, U. 1995. On the Differential Geometry of the Cortical Surface. *Vision Geometry IV*, 2573:304-311.

9. MacDonald, J.D., Kabani, N., Avis, D., and Evans, A.C. 2000. Automated 3-D Extraction of Inner and Outer Surfaces of Cerebral Cortex from MRI. *NeuroImage*, **12**:340–356.
10. Lafferty, J., and Lebanon, G. Diffusion kernels on statistical manifolds. 2004. Technical Report, Department of Computer Science, Carnegie Mellon University.
11. Luders, E., Rex, D.E., Narr, K.L., Woods, R.P., Janke, L., Thompson, P.M., Mazziotta, J.C., Toga, A.W. 2003. Relationships between sulcal asymmetries and corpus callosum size: gender and handedness effects. *Cerebral Cortex* **13**:1084-1093.
12. Rosenberg, S. 1997. *The Laplacian on a Riemannian Manifold*, Cambridge University Press.
13. Sled, J.G., Zijdenbos, A.P., and Evans, A.C. 1988. A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data. *IEEE Transactions on Medical Imaging*, **17**:87-97.
14. Wolfson, E. and Schwartz, E.L. 1989. Computing minimal distances on polyhedral surfaces, *IEEE Trans. on Pattern anal. and Mach. Intel.* **11**:1001-1005.
15. Worsley, K.J., 1994. Local maxima and the expected Euler characteristic of excursion sets of  $\chi^2$ ,  $F$  and  $t$  fields. *Advances in Applied Probability*. **26**:13-42.
16. Worsley et al. 1992. A Three-Dimensional Statistical Analysis for CBF Activation Studies in Human Brain, *Journal of Cerebral Blood Flow and Metabolism* **12**:900-918.
17. Zijdenbos, A.P., Jimenez, A. and Evans. A.C. 1998. Pipelines: Large scale automatic analysis of 3D brain data sets. *NeuroImage*, **7S**:783.