Heat Kernel Smoothing on Human Cortex Extracted from Magnetic Resonance Images

Moo K. Chung

Department of Statistics

Department of Biostatistics and Medical Informatics

University of Wisconsin-Madison

WI 53706., U.S.A.

February 6, 2005

SUMMARY. Gaussian kernel smoothing has been widely used in 3D whole brain imaging analysis as a way to increase signal-to-noise ratio. Gaussian kernel is isotropic in Euclidian space. However, data obtained on the convoluted brain cortex fails to be isotropic in the Euclidean sense. On the curved surface, a straight line between two points is not the shortest distance so one may incorrectly assign less weights to closer observations. In this paper, we will present how to correctly formulate isotropic smoothing for data on the human cortical surface that was extracted from the magnetic resonance images.

As an illustration, we show how to detect the regions of abnormal cortical pattern in 16 autistic children that utilizes the new technique.

KEY WORDS: Heat kernel, Diffusion, Cortex, Laplace Beltrami operator

---

email: mchung@stat.wisc.edu

1. Introduction

The human cerebral cortex has the topology of a 2D highly convoluted grey matter shell of average thickness of 3mm. The thickness of the grey matter shell is usually referred as the *cortical thickness* and can be obtained *in vivo* from magnetic resonance images (MRI). Different clinical populations show different patterns of cortical thickness variation across the cortex. So the cortical thickness can be used an useful index for characterizing cortical shape variations. The thickness measures are obtained after a sequence of image processing steps which are described briefly here. The first step is to classify each voxel into three different tissue types: cerebrospinal fluid (CSF), grey matter and white matter. A neural network classifier (Kollakian, 1996) or a Bayesian mixture modeling (Ashburner and Friston, 2000) have been used for the classification. The CSF/grey matter interface is called the *outer cortical surface* while the grey/white matter interface is called the *inner cortical surface*. These two surfaces bounds the gray matter (Figure 1). Although it is possible to represent the surfaces smoothly and continuously via a thin-plate spline (Xie et al., 2005), so far, the mainstream approach in representing the cortical surface has been to use a fine triangular mesh that is constructed from deformable surface algorithms (MacDonald et al., 2000). Once we have the two surface meshes, we can compute the cortical thickness by computing the distance between the two surfaces (MacDonald et al., 2000). Figure 1 shows an original MRI and its segmentation result based on the neural network classifier. The interface between different tissue types was constructed using the thin-plate spline. The arrow indicates the thickness of the gray matter shell. Figure 2 shows the outer cortical surface mesh generated from a deformable surface algorithm. It consists of 40,962 vertices and 81,920 triangles.

[Figure 1 about here.]

[Figure 2 about here.]

In order to compare cortical thickness measures across subjects, it is necessary to align the cortical surfaces via a surface registration. The concept of surface registration is similar to a curve registration in the functional data analysis (Ramsay and Silverman, 1997). The surface registration tries to register two functional data on a unit sphere (Thompson and Toga, 1996; Robbins, 2003). First a mapping from a cortical surface onto the sphere is established while recording the mapping. Then cortical curvatures are mapped onto the sphere. The two curvature functions on the sphere are aligned by solving a regularization problem that tries to minimize the discrepancy between two functions while maximizing the smoothness of the alignment. This alignment is projected back to the original surface using the recorded mapping. For cross-comparison between subjects, surfaces are registered into a so called *template surface* which serves as a reference coordinates.

The image segmentation, thickness computation and surface registration procedures are expected to introduce noise in the thickness measure. In order to increase the signal-to-noise ratio (SNR) and some type of data smoothing is necessary (Kiebel et al., 1999). For 3D whole brain MRIs, Gaussian kernel smoothing is widely used to smooth data, in part, due to its simplicity in numerical implementation. The Gaussian kernel weights an observation according to its Euclidean distance. However, data residing on the convoluted brain surface fails to be isotropic in the Euclidean sense. On the curved surface, a straight line between two points is not the shortest distance so one may incorrectly assign less weights to closer observations. So when the observations lie on the cortical surface, it is more natural to assign the weight based on the geodesic distance along the surface.

Previously *diffusion smoothing* has been developed for smoothing data along the cortex (Andrade et al., 2001;Chung et al., 2003;Chung and Taylor, 2004). The technique of diffusion smoothing relies on the fact that the Gaussian kernel smoothing in $\mathbb{R}^d$ is equivalent to solving an isotropic diffusion equation in $\mathbb{R}^d$ (Chaudhuri and Marron, 2000). For an overview of using diffusion equations in statistical literature, one may refer to Chaudhuri and Marron (2000) and Ramsay (2000). By solving a diffusion equation on a curved manifold $\partial\Omega$, Gaussian kernel smoothing can be indirectly generalized to $\partial\Omega$. The drawback of this method is the need for estimating the Laplace-Beltrami operator and setting of up a finite element method (FEM) to solve the diffusion equation numerically (Chung and Taylor, 2004). To address this shortcomings, we have developed a simpler method based on the heat kernel convolution on a manifold.

As an illustration, the method was applied to groups of autistic and normal subjects, and we were able to detect the regions of statistically significant cortical thickness difference between the groups.

2. Heat kernel smoothing

The cortical surface $\partial\Omega$ can be assumed to be a C^2 Riemannian manifold (Joshi et al., 1995). Let $p = X(u^1, u^2) \in \partial\Omega$ be the parametric representation of $\partial\Omega$. We assume the following model on thickness measure Y:

$$Y(p) = \theta(p) + \epsilon(p),$$

where $\theta(p)$ is a mean thickness function and $\epsilon(p)$ is a zero-mean random field with covariance function $R_\epsilon(p,q)$. The Laplace-Beltrami operator Δ corresponding to the surface parameterization $p = X(u^1, u^2) \in \partial\Omega$ can be written as

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

where $g = (g_{ij})$ is the Riemannian metric tensor given by the bilinear form $g_{ij} = \langle \frac{\partial X}{\partial u^i}, \frac{\partial X}{\partial u^j}\rangle$. In the case when the metric is flat, i.e. $g = \delta_{ij}$, the Laplace-Beltrami operator becomes the Euclidean Laplacian $\Delta = \frac{\partial^2}{\partial (u^1)^2} + \frac{\partial^2}{\partial (u^2)^2}$ in $\mathbb{R}^2$. By solving equation

$$\Delta\psi = \lambda\psi \tag{1}$$

on $\partial\Omega$, we can find ordered eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots$ and corresponding eigenfunctions $\psi_0, \psi_1, \psi_2, \cdots$. The eigenfunctions ψ_j form orthonormal basis on $L^2(\partial\Omega)$. On a unit sphere, the eigenvalues are $m(m+n-1)$ and the corresponding eigenfunctions are spherical harmonics Y_{lm} (Wahba, 1990). On an arbitrary surface, the explicit representation of eigenvalues and eigenfunction are only obtained through numerical methods. Based on orthonormal basis, the *heat kernel* $K_\sigma(p,q)$ is analytically given as

$$K_\sigma(p,q) = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} \psi_j(p)\psi_j(q), \tag{2}$$

where σ is the bandwidth of the kernel (Rosenberg, 1997; Berline et al., 1991). When $g_{ij} = \delta_{ij}$, the heat kernel becomes a Gaussian kernel

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

Hence the heat kernel is a natural extension of the Gaussian kernel. Under some regularity condition, K_σ is a probability distribution on $\partial\Omega$ so that $\int_{\partial\Omega} K_\sigma(p,q)\ dq = 1$. This can be interpreted as the transition probability density for an isotropic diffusion process with respect to Riemannian volume element (Wang, 1997). The kernel is symmetric, i.e. $K_\sigma(p,q) = K_\sigma(q,p)$ and isotropic. The property of a kernel being isotropic needs some explanation. Let us first define the geodesic distance on a curved surface. Consider a curve segment $C \subset \partial\Omega$ that connects p and q, and parameterized by $\gamma_C(t)$ with $\gamma_C(0) = p$ and $\gamma_C(1) = q$. In Cartesian coordinates, $\gamma_C(t) = (\gamma_C^1(t), \cdots, \gamma_C^d(t)) \in \mathbb{R}^d$. The arclength of C is given by

$$\int_0^1 \langle \frac{d\gamma_C}{dt}, \frac{d\gamma_C}{dt} \rangle^{1/2}\ dt = \int_0^1 \Big[\sum_{i,j} g_{ij} \frac{d\gamma_C^i}{dt} \frac{d\gamma_C^j}{dt} \Big]^{1/2}\ dt.$$

Then the geodesic curve connecting p and q is defined as a minimizer

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

Function f is *isotropic* on surface $\partial\Omega$ if $f(p) =$ *constant* for all point p on geodesic circle $d(0,p) =$ constant. Since $K_\sigma(p,q)$ has two arguments while symmetric, the isotropic property holds for either one of the arguments.

Now we define *heat kernel smoothing* of cortical thickness Y to be the convolution:

$$K_\sigma * Y(p) = \int_{\partial\Omega} K_\sigma(p,q) Y(q)\ dq. \tag{3}$$

Let us list a couple of important properties of heat kernel smoothing.

Property. (1) $K_\sigma * Y$ is the unique solution of the following isotropic diffusion equation at time $t = \sigma^2/2$:

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

This is a well known result (Rosenberg, 1997). The previous diffusion smoothing approach smooth data by directly solving the diffusion equation (Chung and Taylor, 2004). This also shows that the heat kernel smoothing isotropically assigns weights along $\partial\Omega$.

Property. (2) $K_\sigma * Y(p) = \arg\min_{\theta(p)\in L^2(\partial\Omega)} \int_{\partial\Omega} K_\sigma(p,q)\big[Y(q) - \theta(p)\big]^2\, dq.$

Convolution (3) can be viewed in the context of regularization on a manifold by this property. This generalizes a similar discrete result given for Gaussian kernel smoothing in the Euclidean space (Fan and Gijbels, 1996). It can be proved easily noting that the integral can be written as quadratic in θ: $K_\sigma * Y^2(p) - 2\theta K_\sigma * Y(p) + \theta^2$.

It is natural to assume two random fields $\epsilon(p)$ and $\epsilon(q)$ to have less correlation when p and q are away. So we assume ρ to be non increasing. Suppose we have a isotropic covariance function of type $R_\epsilon(p,q) = \rho(d(p,q))$ for some nondecreasing function ρ. Then we can show the variance reduction property of heat kernel smoothing.

Property. (3) $\mathbf{Var}[K_\sigma * Y(p)] \le \mathbf{Var}Y(p)$ for each $p \in \partial\Omega$.

See Appendix A for the proof. It is believed that the requirement for the covariance function may be relaxed.

Property. (4) $\lim_{\sigma\to 0} K_\sigma * Y = Y.$

This can be easily seen from the fact that as $\sigma \to 0$, the heat kernel becomes the Dirac delta function.

Property. (5)

$$\lim_{\sigma \to \infty} K_\sigma * Y = \frac{\int_{\partial\Omega} Y(q)\, dq}{\mu(\partial\Omega)},$$

where $\mu(\partial\Omega)$ is the surface area of $\partial\Omega$. The property 5 shows that when we choose large bandwidth, heat kernel smoothing converges to the sample mean of data on $\partial\Omega$. This is easily proved by noting

$$\widetilde{K}_\sigma * Y(p) = \frac{\int_{B_p} \exp\big[-\frac{d^2(p,q)}{2\sigma^2}\big] Y(p)\, dq}{\int_{B_p} \exp\big[-\frac{d^2(p,q)}{2\sigma^2}\big]\, dq} \to \frac{\int_{B_p} Y(q)\, dq}{\mu(B_p)}$$

as $\sigma \to \infty$. Now by letting B_p to cover the whole cortex $\partial\Omega$, we prove the property. A similar result in the context of differential geometry can be found in Rosenberg, 1997.

Property. (6)

$$\underbrace{K_\sigma * \cdots * K_\sigma}_{k \text{ times}} * Y = K_{\sqrt{k}\sigma} * Y.$$

This can be seen as the scale space property of diffusion. From Property (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 σ^2. Hence

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

Arguing inductively we see that the general statement holds. We will note the k-fold iterated kernel as $K_\sigma^{(k)} = \underbrace{K_\sigma * \cdots * K_\sigma}_{k \text{ times}}$.

The problem with the heat kernel is that it is almost impractical to determine the eigenvalues and eigenfunction os the Laplace-Beltrami operator in an arbitrary surface like the human brain cortex. To address this problem we use the *parametrix expansion* of the heat kernel (Rosenberg, 1997; Wang, 1997):

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

for small $d(p,q)$. This expansion spells out the exact form of the kernel for small bandwidth. When the metric is flat, the heat kernel becomes a Gaussian kernel, reconfirming that heat convolution is a generalization of Gaussian kernel. The expansion is the basis of our heat kernel smoothing formulation. Heat kernel smoothing with a large bandwidth will be decomposed into iterated kernel smoothing.

We will truncate and normalize the heat kernel using the first order term. For each p, we define

$$\widetilde{K}_\sigma(p,q) = \frac{\exp\big[-\frac{d^2(p,q)}{2\sigma^2}\big]\mathbf{1}_{B_p}(q)}{\int_{B_p}\exp\big[-\frac{d^2(p,q)}{2\sigma^2}\big]\,dq}, \tag{6}$$

where $\mathbf{1}_{B_p}$ is an indicator function defined on a small compact domain containing B such that $\mathbf{1}_{B_p}(q) = 1$ if $q \in B_p$ and $\mathbf{1}_{B_p}(q) = 0$ otherwise. Note that for each fixed p, $\widetilde{K}_\sigma(p,q)$ defines a probability in B_p and it converges to $K_\sigma(p,q)$ as $\sigma \to 0$ in B_p This implies

$$\widetilde{K}_\sigma^{(k)} * Y(p) \to K_\sigma^{(k)} * Y(p) \text{ as } \sigma \to 0.$$

We can proved this only when $g_{ij} = \sigma_{ij}$ since the exact analytical expression of the heat kernel is nonexistent and still an ongoing research problem (Rosenberg, 1997; Wang, 1997).

Property. (7) When $g_{ij} = \delta_{ij}$,

$$K_{\sqrt{k}\sigma} * Y(p) \le \widetilde{K}_\sigma^{(k)} * Y(p) \le \frac{1}{\alpha^k(p)} K_{\sqrt{k}\sigma} * Y(p),$$

where $\alpha(p) = \int_{B_p} K_\sigma(p,q)\,dq$.

For the proof, see Appendix B. The upper bound is valid without the assumption of the flat metric. As $\sigma \to 0$, $\alpha(B_p) \to 1$ and the inequalities collapse proving the convergence. For sufficiently small σ and B_p, we can make $\alpha(B_p)$ as small as possible. Hence, decreasing the size of bandwidth and increasing the number of iterations will perform better although it will be slow down the algorithm.

For a discrete triangular mesh, we can take B_p to be a set of points containing p and its neighboring nodes $q_1, \cdots, q_m$, and take a discrete measure on B_p in the integral, which will make 6 still a probability distribution. This can be viewed as a *Gaussian kernel Nadaraya-Watson* type smoothing extended to manifolds (Fan and Gijbels, 1996; Chaudhuri and Marron, 2000). Figure 3 shows the heat kernel smoothing on both a single subject data and simulated data. The cortical thickness measures are projected onto a template that has less surface folding to show the progress of smoothing. The bottom figure shows the progress of heat kernel smoothing with $\sigma = 1$ and upto $k = 5000$ iterations applied to a simulated data. From 12 normal subjects, the mean thickness $\theta(p)$ and variance $R_\epsilon(p,p)$ functions are estimated and used to generate random fields with a Gaussian white noise. It begin to show the convergence to the sample mean thickness over all cortex.

[Figure 3 about here.]

3. Multiple comparisons on the human cortex

We will briefly describe how to perform multiple comparisons on $\partial\Omega$ using the random field theory. The random field theory based approach is widely used for correcting multiple comparisons in brain imaging. We let the first group to be autistic and the second group to be normal control. There are n_i

subjects in the i-th group. For the i-th group, we have the following model on cortical thickness Y_{i_j} for i-th group and j-th subject:

$$K_\sigma * Y_{i_j}(p) = \theta_i(p) + \epsilon_{i_j}(p),$$

where θ_i is the mean thickness of the i-th group and ϵ_{i_j} is independent zero mean smooth Gaussian random fields. We assume the noise to be a Gaussian white noise convolved with heat kernel K_σ, i.e. $\epsilon_{i_j} = K_\sigma * W$. Then we test if the mean thicknesses for two groups are the same at every points, 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 null hypothesis is the intersection of collection of hypothesis

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

where $H_0(p) : \theta_1(p) = \theta_2(p)$. for fixed p. The test statistic is given by

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

where the pooled variance $S^2 = ((n_1-1)S_1^2 + (n_2-1)S_2^2)/(n_1+n_2-2)$. Under H_0, it is the t random field with $n = n_1 + n_2 - 2$ degrees of freedom (Worsley, 1994). The type I error for the multiple comparisons is given by

$$\begin{aligned}
\alpha &= P\Big(\bigcup_{p\in\partial\Omega}\{T(p) > h\}\Big) = 1 - P\Big(\bigcap_{p\in\partial\Omega} T(p) \le h\}\Big) \\
&= 1 - P(\sup_{p\in\partial\Omega} T(p) \le h) = P(\sup_{p\in\partial\Omega} T(p) > h)
\end{aligned}$$

for some h. The resulting p-value is usually called the *corrected p-value* in brain imaging. The distribution of $\sup_{p\in\partial\Omega} T(p)$ is asymptotically given as

$$P(\sup_{p\in\partial\Omega} T(p) > h) \approx \sum_{d=0}^{2} \phi_d(\partial\Omega)\rho_d(h) \tag{7}$$

where ϕ_d are the d-dimensional Minkowski functionals of $\partial\Omega$ and ρ_d are the d-dimensional Euler characteristic (EC) density of t-field (Worsley, 2003). The Minkowski functionals are $\phi_0 = 2, \phi_1 = 0, \phi_2 = \text{area}(\partial\Omega)/2 = 49,616\text{mm}^2$, the half area of the template cortex $\partial\Omega$. The EC density is given by

$$\begin{aligned}
\rho_0(h) &= \int_h^\infty \frac{\Gamma(\frac{n+1}{2})}{(n\pi)^{1/2}\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{d}\right)^{-\frac{(n+1)}{2}} dx, \\
\rho_2(h) &= \frac{\lambda}{(2\pi)^{3/2}}\frac{\Gamma(\frac{n+1}{2})}{(\frac{n}{2})^{1/2}\Gamma(\frac{n}{2})}\left(1+\frac{h^2}{d}\right)^{-\frac{(n-1)}{2}} h,
\end{aligned}$$

where λ measures the smoothness of field $\epsilon = \epsilon_{i_j}$ and given as $\lambda = 1/(2\sigma^2)$.

If we want to removed the effect of age and total grey matter volume, we set up a general linear model (GLM) on cortical thickness Y_j for subject j

$$Y_j(p) = \lambda_1(p) + \lambda_2(p)\cdot \texttt{age}_j + \lambda_3(p)\cdot \texttt{volume}_j + \beta(p)\cdot \texttt{group}_j + \epsilon_j \tag{8}$$

where dummy variable `group` is 1 for the autistic subjects and 0 for the normal subjects. Traditionally GLM has been a very popular approach in brain imaging (Friston., 2002). Then we test the group difference

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

v.s.

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

The test statistic is the ratio of the sum of the squared residual errors and under H_0, it is a F random random field with 1 and $n = n_1+n_2-4$ degrees of freedom. F fields has a similar asymptotic results like (7). The EC-densities of F fields are given in Worsley (1994). The resulting corrected p-values maps for both t and F tests are shown in Figure 4. The main use of the corrected p-value maps are the localization and visualization of signal difference.

[Figure 4 about here.]

4. Application

Gender and handedness affect brain anatomy so all the 16 autistic and 12 control subjects used in the study were screened to be right-handed males. Sixteen autistic subjects were diagnosed with autism. The average age for the control subject is 17.1 ± 2.8 and the autistic subjects is 16.1 ± 4.5. High resolution anatomical magnetic resonance images (MRI) were obtained using a 3-Tesla GE SIGNA scanner. Afterwards, MRIs are obtained and both the outer and inner cortical surfaces are extracted via a deformable surface algorithm (MacDonald et al., 2000). The resulting triangular meshes consist of 40,962 vertices and 81,920 triangles with the average edge length of 3 mm (Figure 2). The cortical distance is computed between two surfaces following the method in MacDonald et al. (2000). The thickness measures are smoothed with heat kernel smoothing with parameters $\sigma = 1$ and $k = 200$ giving the effective smoothness of $\sqrt{200} = 14.14$. A surface-to-surface registration to a template surface was performed following the curvature matching method of Robbins (2003). Then following the multiple comparisons procedure of the previous section, the corrected p-value maps for the both t and F statistics results are projected onto the template surface. Figure 4 shows statistically significant regions of cortical thickness between two groups. The left two images are the corrected p-value map (< 0.1) showing mainly thinner cortical shell in autism. The upper (lower) scale shows the regions of thicker (thinner) cortical shell in autism. However, most of these regions turned out to be not significant after removing the effect of age and total grey matter volume via a F random field. The two right images are the corrected p-value map (< 0.1) of the F statistic result. After removing the effect of age and grey matter, the statistically significant regions of thickness decrease are highly localized at the right inferior orbital prefrontal cortex, the left superior temporal sulcus and the left occipito-temporal gyrus.

5. Discussion

There are many aspects of heat kernel smoothing that we have ignored in this paper. The purpose of this paper is to introduce heat kernel smoothing and its use in brain imaging to attract more attention of researchers on the important new problem. Although splines on a unit sphere has received a great deal of attention (Freeden, 1981; Wahba, 1990), the idea of smoothing or regularizing data on arbitrary manifolds has not been investigated by many researchers possibly due to the fact that there are no manipulatable basis functions available. The Property (2) gives a basic framework for a regression on a manifolds. Instead of trying to explicitly determine the coefficient of the mean function

$$\theta(p) = \sum_{j=0}^{n} c_j \psi_j(p)$$

that minimizes the cost function in the Property (2), heat kernel smoothing determines them implicitly. In fact, c_j's are given by

$$c_j = e^{-\lambda_j \sigma} \int_{\partial\Omega} \psi_j(q) Y(q) \, dq.$$

However, eigenvalues λ_j and eigenfunctions ψ_j can not be determined analytically other than algebraic surfaces like a sphere or a torus. Our method avoids trying to determine c_j's explicitly and solve it for θ that is a solution to an isotropic diffusion equation. So our approach works on non algebraic surfaces like the human brain cortex. Recently researchers are begin to investigate the problem of regularization on manifolds or graphs in the context of classification of massive data (Belkin, 2003; Kondor and Lafferty., 2002). In a computer vision area, diffusion equations have been also used as a way to smooth noisy manifolds itself (Bulow, 2002) and researchers are begin to diffuse signals on manifolds as well. As noted by Chaudhuri and Marron (2000), there is a lack of using diffusion equations in statistical literature. It is hoped that this paper will provide a motivation for developing new methodologies for smoothing measurements on surfaces.

We have implemented our heat kernel smoothing in MATLAB and it is freely available with a sample cortical thickness data for research community. They can be found at
`http://www.stat.wisc.edu/~mchung/softwares/hk/hk.html`. The following codes will load data into MATLAB and smooth data with parameters $\sigma = 1$ and $k = 200$.

```
[tri,coord,nbr,normal]=mni_getmesh('outersurface.obj');
load thickness.data;
output=hk_smooth(thickness',tri,coord,nbr,1,200);
trisurf(tri,coord(1,:),coord(2,:),coord(3,:),output);
```

Acknowledgements

The author wish to thank Kim Dalton and Richard J. Davidson of the Waisman Laboratory for Brain and Behavior for providing the autism data and Steve Robbins and Alan Evans of Montreal Neurological Institute for providing technical assistant with surface registration.

References

Andrade, A., Kherif, F., Mangin, J., Worsley, K., 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.

Ashburner, J. and Friston, K. (2000). Voxel-based morphometry - the methods. *NeuroImage* **11**, 805–821.

Belkin, M., N. P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* **15**, 1373–1396.

Berline, N., Getzler, E. and M., V. (1991). *Heat kernels and dirac operators.* Springer-Verlag.

Bulow, T. (2002). Spherical diffusion for 3d surface smoothing. *1st Int'l Symp. on 3D Data Proc. Vis. and Transm.* .

Chaudhuri, P. and Marron, J. S. (2000). Scale space view of curve estimation. *The Annals of Statistics* **28**, 408–428.

Chung, M. and Taylor, J. (2004). Diffusion smoothing on brain surface via finite element method. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI).*

Chung, M., Worsley, K., Robbins, S. and Evans, A. (2003). Tensor-based brain surface modeling and analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 467–473.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC.

Freeden, W. (1981). O spherical splin interpolation and approximation. *Math. Meth. Appl. Sci.* **3**, 551–575.

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

Joshi, S., Wang, J., Miller, M., Van Essen, D. and Grenander, U. (1995). On the differential geometry of the cortical surface. *Vision Geometry IV* pages 304–311.

Kiebel, S., Poline, J.-P., Friston, K., Holmes, A. and Worsley, K. (1999). Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *NeuroImage* **10**, 756–766.

Kollakian, K. (1996). Performance analysis of automatic techniques for tissue classification in magnetic resonance images of the human brain. Technical Report Master's thesis, Concordia University, Montreal, Quebec, Canada.

Kondor, I. and Lafferty., J. (2002). Diffusion kernels on graphs and other discrete structures. In *Proceedings of ICML*.

MacDonald, J., Kabani, N., Avis, D. and Evans, A. (2000). Automated 3-d extraction of inner and outer surfaces of cerebral cortex from mri. *NeuroImage* **12**, 340–356.

Ramsay, J. (2000). Differential equation models for satistical functions. *The*

Canadian Journal of Statistics **28**, 225–240.

Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis.* Springer.

Robbins, S. (2003). 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.

Rosenberg, S. (1997). *The Laplacian on a Riemannian Manifold.* Cambridge University Press.

Thompson, P. and Toga, A. (1996). A surface-based technique for warping 3-dimensional images of the brain. *IEEE Transactions on Medical Imaging* **15**, 1–16.

Wahba, G. (1990). *Spline models for observational data.* SIAM.

Wang, F.-Y. (1997). Sharp explict lower bounds of heat kernels. *Annals of Probability* **24**, 1995–2006.

Worsley, K. (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.

Worsley, K. (2003). Detecting activation in fmri data. *Statistical Methods in Medical Research.* **12**, 401–418.

Xie, X., Chung, M. and Grace, W. (2005). Magnetic resonance image segmentation with thin plate spline thresholding. In *11th Annual Meeting of the Organization for Human Brain Mapping.* Toronto, submitted.

Appendix A

Proof of Property (3)

Note that

$$R_Y(p', q') = R(p', q') = \rho(d(p', q')) \leq \rho(0) = \mathbf{Var}Y(p').$$

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

$$\begin{aligned} R(p,q) &= \mathbb{E}\Big[\int_{\partial\Omega} K_\sigma(p,p')Y(p')\,dp' \int_{\partial\Omega} K_\sigma(q,q')Y(q')\,dq'\Big] \\ &= \int_{\partial\Omega}\int_{\partial\Omega} K_\sigma(p,p')K_\sigma(q,q')R_Y(p',q')\,dp'\,dq' \\ &\leq \int_{\partial\Omega}\int_{\partial\Omega} K_\sigma(p,p')K_\sigma(q,q')\rho(0)\,dp'\,dq' \\ &= \rho(0) \end{aligned}$$

since K_σ is a probability distribution. Now letting $p = q$, we have

$$\mathbf{Var}[K_\sigma * Y(p)] = R(p,p) \leq \mathbf{Var}Y(p).$$

Appendix B

Proof of Property (7)

We only prove for $k = 2$ and the result follows inductively. The heat kernel can be written as

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

Applying convolution again, we have

$$K_\sigma * K_\sigma(p,q) \;\geq\; \alpha K_\sigma * \widetilde{K}_\sigma(p,q) \geq \alpha^2 \widetilde{K}_\sigma * \widetilde{K}_\sigma(p,q).$$

Now let us find a lower bound. When the metric is flat, $d(p,q)$ is the Euclidean distance and K_σ is the usual Gaussian kernel. Hence, $K_\sigma(p,q) \leq \widetilde{K}_\sigma(p,q)$ for $q \in B_p$. Then $K_\sigma * K_\sigma(p,q) \leq \widetilde{K}_\sigma * \widetilde{K}_\sigma(p,q)$. Now apply $\widetilde{K}_\sigma$ to signal Y which proves the statement for $k = 2$.

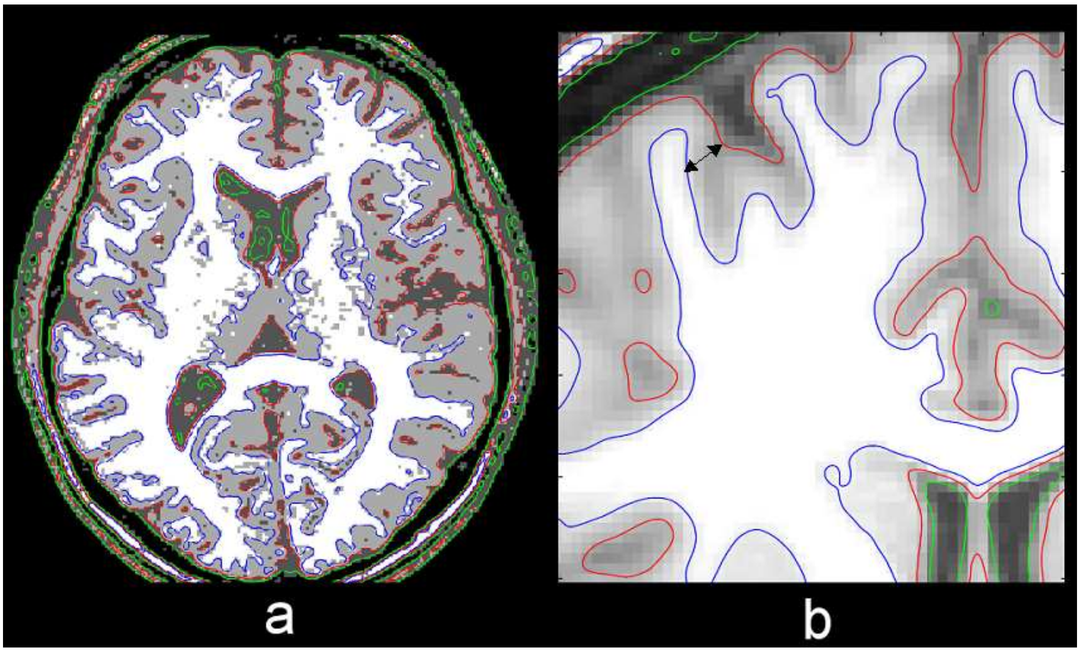

Figure 1. Left: segmented magnetic resonance image of brain. The gray matter is in fact gray in color. Right: Enlarged image showing the outer and inner surface. The arrow indicates the cortical thickness between two surfaces. The boundaries are generated using a thin-plate spline (Xie et al., 2005).

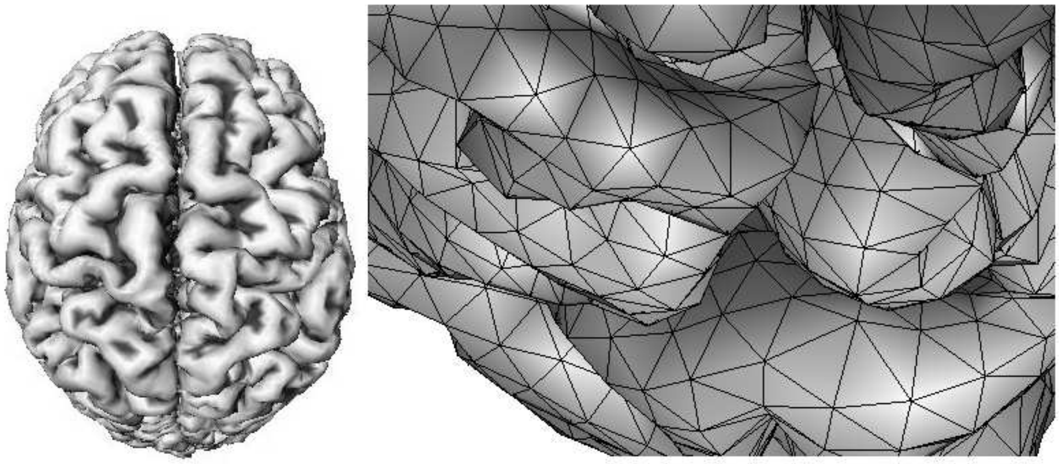

Figure 2. Left: Outer cortical surface. Right: A close up image of the outer cortical surface showing connected triangle elements. The cortical thickness measures are obtained at each node.

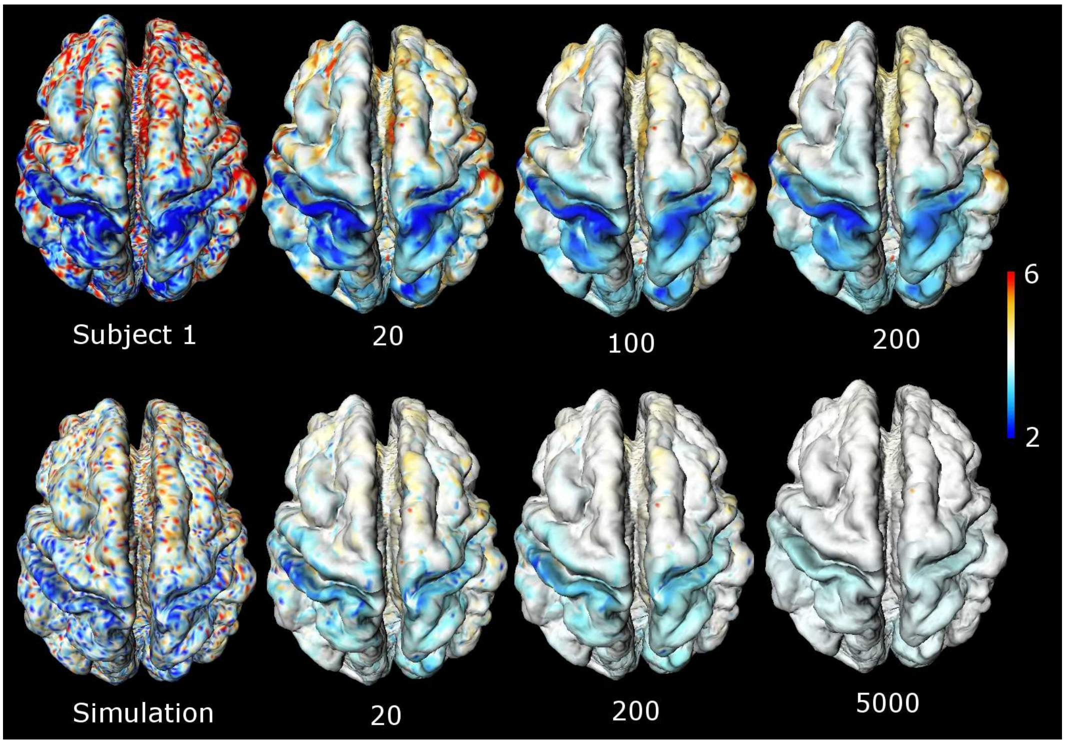

Figure 3. Top: Heat kernel smoothing of the first subject in the control group with $\sigma = 1$ and $k = 20, 100, 200$ iterations. The thickness measures are displayed between 2 and 6 mm. Bottom: Heat kernel smoothing on simulated data. Mean thickness function $\theta(p)$ and variance are estimated from 12 normal subject data and Gaussian white noise is added to the mean function. Parameters are $\sigma = 1, k = 20, 200, 5000$.

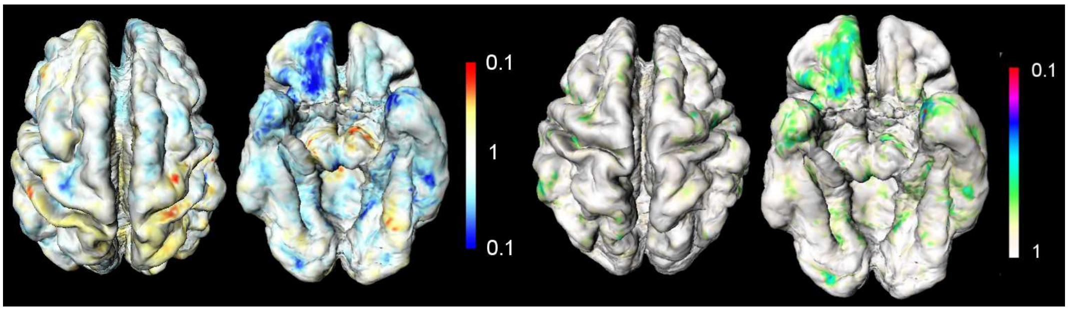

Figure 4. Corrected p-value maps projected onto the template. Left: p value of t statistic map. Right: p-value of the F statistic map. Comparing two p-value maps, we conclude that the thicker grey matter region is largely due to the effect of age and grey matter volume difference.