Project: Heat kernel construction on cortical surface using geodesic distance

STAT 992: Statistical Methods in Signal and Image Analysis

Instructor: Moo K. Chung

Tulaya Limpiti

May 15, 2004

1 Introduction

In medical imaging, measured data always suffer from noise. One way to effectively increase the signal-to-noise ratio is to perform smoothing on the data, where data points are weighted-averaged with their neighbors. Smoothing can be viewed as low-pass filtering, since the high-frequency components are removed from the smoothed data. In brain imaging, the data points lies on the cortical surface, and the underlying functional data are naturally smooth, for example, cortical thickness, blood oxygenation level dependent, or neural activity patterns [7,1,2]. To estimate signal component from noisy observations, kernel-based smoothing is the most popular non-parametric estimator. Different kernels have been used in the smoothing process, such as uniform (square) kernel, Gaussian kernel, Epanechnikov kernel, and triweight kernel [16]. The most widely used kernel is Gaussian kernel, which is originally defined on an n-dimensional Euclidean space.

Nevertheless, in brain imaging our data of interest lies instead on a cortical surface, which is assumed to be a smooth two-dimensional Riemannian manifold [8]. Because of the convoluted nature of the cortical surface, two points on the surface that appear close together in a Euclidean sense can actually be very far apart, for example, when each of the two points lies on an opposite bank of a sulcus (see figure 1). Therefore, if we were to apply Gaussian kernel smoothing on this manifold using a kernel that is isotropic in Euclidean space, we may give large weights to data that are far away, and small weights to data that are closer. In other words, we would mistakenly put higher emphasis on the less correlated data in the resulting weighted average, which would be inappropriate. An alternative smoothing method is diffusion smoothing using diffusion (heat) kernel. It has been shown [6,8,11] that there is a tight relationship between heat kernel and Gaussian kernel. On geometric manifolds, heat kernel is a function of geodesic distance—shortest distance along the surface—instead of Euclidean distance. Geodesic distance provides a proper measure of correlations between data points.

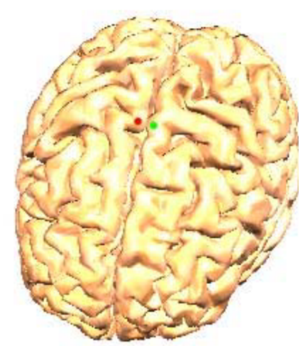

Figure 1: Cortical surface. The red and green dots are close in the Euclidean sense. But to get from one dot to the other requires a long traveling distance along the curve (down the sulcus and back up).

In this report, we construct heat kernels based on geodesic distances of data points on the cortical surface and perform kernel smoothing using our kernels on cortical thickness measurement data. We note that our heat kernel is locally equivalent to an isotropic Gaussian kernel, and hence the amount of smoothness can be controlled by adjusting the bandwidth matrix of the kernel. The geodesic distances calculation in our kernel construction is carried out via dynamic programming. The generalization of the result to anisotropic kernels smoothing is possible, although it is beyond the scope of this project.

To represent the cortical surface, we have adopted the polygonal representation, commonly used in the computer vision community. The problem of finding minimal distances between arbitrary points on polyhedral surface has attracted interests of many researchers in the field of graph theory, robotics motion planning, computer-aided neuroanatomy, to name a few. Several algorithms have been developed. The first algorithm is that of Dijkstra [10] where he solved the problem on graphs in 1959. Sharir and Schorr [15] proposed the algorithm that finds the shortest path in $O(n^3 \log n)$, where n is the total number of nodes on the surface. However, they constrained themselves to consider the problem only on convex polyhedrons. Mitchell, Mount, and Papadimitriou came up with an improved algorithm with continuous Dijkstra structure and the computational complexity of $O(n^2 \log n)$ [12]. Other algorithms include those contributed by Chen and Han [5], and Wolfson and Schwartz [17]. Among these algorithms, dynamic programming (DP) finds itself become popular, due to its flexibility and its ease of implementation for discrete constraint sets. The complexity of DP is $O(n^2)$.

This report is organized as followed. In the next section we review some relevant backgrounds necessary to understand our proposed method. We start with the definition of Gaussian kernel and heat kernel and their uses in data smoothing and estimation in section 2.1. Riemannian manifold and its polygonal representation are explained in section 2.2, where as in section 2.3 we give a mathematical description of geodesic distance. Section 2.4 introduces the concepts of dynamic programming. In section 3 we give details of our proposed construction of isotropic heat kernel on the cortical surface using geodesic distance. Experimental results using cortical thickness measurement data are presented in section 4. Finally, we conclude with some discussions on our results in the last section.

Throughout the report, matrix and vector quantities are written in bolded-face uppercase and lowercase letters, respectively. Superscript $\{\cdot\}^T$ denotes matrix transpose, and superscript $\{\cdot\}^{-1}$ denotes matrix inverse.

2 Backgrounds

2.1 Gaussian kernel and Heat kernel

Generally, a *Gaussian kernel* is defined in an n-dimensional Euclidean space as

$$K_H(x) = (2\pi)^{-\frac{n}{2}} (\det HH^T)^{1/2} \exp\left(-\frac{x^T (HH^T)^{-1} x}{2}\right) \tag{1}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T \in \Re^n$. The n-by-n matrix **H** is called the *bandwidth matrix* of the kernel [6]. For isotropic kernel, $\mathbf{H} = \sigma^2 \mathbf{I_n}$, where $\mathbf{I_n}$ is an n-by-n identity matrix and σ^2 is a user-defined positive number. With this bandwidth matrix, the isotropic kernel has the form:

$$K_\sigma(x) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2}\right) \tag{2}$$

The bandwidth matrix governs the amount of smoothing, There is a bias-variance tradeoff between small and large bandwidth. Small bandwidth corresponds to undersmoothing, which has small bias and large variance. On the other hand, oversmoothing, with large bias and small variance, results from large bandwidth selection [13].

In image analysis, we treat observations as a set of samples drawn from a continuous function, corrupted by noise. To separate the signal component from the noise component using isotropic Gaussian kernel smoothing, the kernel is convolved with the signal,

$$\mu(x) = \int_{R^n} K_H(x-y) f(y) dy. \tag{3}$$

Theoretically, Gaussian kernel is desirable because it is continuously differentiable, separable, and symmetric. In practice, care has to be taken in implementing a Gaussian kernel because of its infinite support. We use the discrete, truncated version of the kernel, as given in [6]:

$$\widetilde{K}_H(x_i) = \frac{K_H(x_i) 1_\Omega(x_i)}{\sum_{\forall x_i \in \Omega} K_H(x_i)} \tag{4}$$

where Ω is the constraint region. In nonparametric regression setting, this is called the Nadaraya-Watson estimator [7].

Given observations $Y(\mathbf{x})$, $\mathbf{x} \in M$, *heat kernel* $K_t(\mathbf{x},\mathbf{y})$ on geometric manifolds is governed by the second-order differential equation $\frac{\partial f}{\partial t} = \Delta f$, with the initial condition f(**x**,0) = Y(**x**). Δ is the Laplace-Beltrami operator. The solution of this PDE is

$$f(x,t) = \int_M K_t(x,y) f(y) dy \tag{5}$$

For the simple manifold, M = ℜ, K_t(**x**,**y**) is equivalent to an isotropic Gaussian kernel with bandwidth $\sigma^2 = 2t$ [11]. In other words, the result of the diffusion of initial data Y for the time duration $t = \sigma^2/2$ is similar to the result of the convolution of an isotropic Gaussian kernel ($\sigma^2 = 2t$) with the data Y.

The parametric expansion of the heat kernel has the form [11]:

$$P_t^{(m)}(x,y) = (4\pi t)^{-1/2} \exp\left(-\frac{d^2(x,y)}{4t}\right) \cdot \left[\psi_0(x,y) + \psi_1(x,y)t + \ldots + \psi_m(x,y)t^m\right] \tag{6}$$

Assuming t is sufficiently small and the point **x**, **y** are close together, the heat kernel can be approximated locally as [6]:

$$K_t(x,y) \approx (4\pi t)^{-1/2} \exp\left(-\frac{d^2(x,y)}{4t}\right) \tag{7}$$

It is obvious that the approximation resembles the form of isotropic Gaussian kernel with bandwidth 2t. The only difference is that the Euclidean distance term in the exponent is replaced by geodesic distance. Therefore, geodesic distance turns the Riemannian manifold into a metric space, satisfying the usual properties, e.g., positivity, symmetry, and triangle inequality. Analytically, the heat kernel in (7) is the kernel we attempt to construct, of which the details are given in section 3.

One advantage of the diffusion smoothing is that it solves the problem usually encountered with Gaussian kernel smoothing at the boundary. Fortunately, the cortical surface is a compact manifold, and so the problem at boundaries is not applicable in our specific case.

Another drawback of kernel smoothing that has been pointed out in [16] is the difficulty in finding the appropriate value of the bandwidth to be used. One solution is to plot the integrated squared error (ISE) between the smoothed data and the original signal. The optimal bandwidth is identified as the global minima of the ISE plot. However, the original, noise-free signal is rarely known in advance, which makes the ISE idea only suitable for algorithm validations. Simonoff also suggests several other bandwidth selection methods. One interesting method that exploits the geometry of the manifold is called *local-varying bandwidth*, where distance from point x to its k^{th} nearest neighbor (for some fixed k) is used as σ^2, results in more smoothing in the regions of low density.

In this region, distance from x to the kth neighbor would be large, and so as the bandwidth. Unfortunately this local-varying bandwidth does not give satisfactorily results, and some subtle features of the data are missed.

In practice, an alternative solution is called *iterated kernel smoothing* [6,7]. Argued by induction, an iterated kernel smoothing formula is

$$K_\sigma^{(m)} * f = \underbrace{K_\sigma * K_\sigma * \ldots *}_{m} f = K_{\sqrt{m}\sigma} * f \quad (8)$$

This formula says that smoothing with large bandwidth to be achieved by iteratively apply smoothing using smaller bandwidth. We make use of this interesting property in our experiments in section 4.

2.2 Riemannian manifold and its polygonal representation

A *Riemannian manifold* (M,g) is a differentiable manifold with a family of smoothly varying, positive definite inner products $g = g_p$ defined on a tangent space T_pM for each p $\in$ M [11]. The *Riemannian metric tensors* **G** = $\{g\}_{ij}$ is given as

$$G_p(v,w) = \sum_{i,j} g_{ij}(p) v_i w_j$$, where **v** and **w** belong to T_pM.

More Pages