Moo K. Chung<sup>\*</sup> and Yanli Wang<sup>†</sup>

\* Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI 53706, USA mkchung@wisc.edu

<sup>†</sup>Institute of Applied Physics and Computational Mathematics, P. O. Box 8009, Beijing 100088, P. R. China wangyanliwyl@gmail.com

We review the heat kernel smoothing techniques for denoising and regressing data in irregularly shaped domains embedded in Euclidean spaces. This is a problem often encountered in functional data analysis and medical imaging. In this chapter, we present a unified mathematical framework based on the eigenfunctions of the Laplace-Beltrami operators defined on irregular domains. Numerical implementation issues will be addressed as well. Various examples will be presented. We also present a few new theoretical results on the properties of heat kernel smoothing.

#### Contents

1	Introduction	182
<b>2</b>	Laplace-Beltrami eigenfunctions	183
	2.1 Graph Laplacian	184
	2.2 Laplacian in positive definite symmetric matrices	186
3	Heat kernel	188
	3.1 Heat kernel on spheres	188
	3.2 Heat kernel on graphs	189
4	Heat kernel smoothing	190
	4.1 Asymptotics	193
	4.2 Inequalities	194
	4.3 Discrete heat kernel smoothing on graphs	198
5	Statistical properties of heat kernel smoothing	199
	5.1 Heat kernel regression on manifolds	201
	5.2 Statistical properties on graphs	204

#### M. K. Chung & Y. Wang

	5.3 Persistent homology in heat kernel smoothing	205
6	Discussion	205
Re	eferences	206

## 1. Introduction

For *irregular domains* often encountered in images, boundary shapes are often very complex. This causes the geometric shape of the boundary to strongly affect the shape of Gaussian kernels. In such irregular domains, the use of Gaussian kernels may not be appropriate. So there is need to incorporate the shape of the boundary into the shape of kernels. The traditional methods include domain embedding methods which embed the domain of interest within a 2D rectangle or 3D box, where the usual sine and cosine basis are known [8]. Such methods still introduce the ringing artifacts (Gibbs phenomenon) along the boundary [12].

Heat kernel has been popular in shape modeling in recent years. Heat kernel is often used as a natural generalization of the Gaussian kernel. [3, 4] used the truncated Gaussian kernel in locally approximating the heat kernel in manifold learning. [45] used the heat kernel as a multiscale shape feature for surface meshes. [6] used the heat kernel signature (HKS), which is the trace of heat kernel, as an isometry-invariant multi-scale shape descriptor in computer vision. [31] computed the heat kernel on graphs using graph Laplacian for face representation. Also there have been significant developments in kernel methods in machine learning [39, 34, 41, 44, 51]. Most kernel methods in machine learning deal with the linear combination of kernels as a solution to penalized regressions. In most applications, the discrete versions of heat kernel, which in turn uses the eigenvector of graph Laplacian, are often used [42]. The connection between the eigenfunctions of continuous and discrete Laplacians has been well established by several studies [23, 46].

Heat kernel smoothing was introduced in [17, 16] to filter out noisy cortical thickness defined on brain surface mesh vertices obtained from magnetic resonance images [17, 16] approximates the heat kernel locally by iteratively applying Gaussian kernel with smaller bandwidth. For recent spectral formulation to heat kernel smoothing [40, 15] constructs the heat kernel analytically using the eigenfunctions of the Laplace-Beltrami (LB) operator, avoiding the need for the linear approximation using Gaussian kernel.

We will start by reviewing the basic spectral geometry related to heat kernel.





Fig. 1. Six representative eigenfunctions on human amygdala surface. The number j represents eigenfunction  $\psi_j$ . The eigenfunctions of the Laplace-Beltrami operators are computed using the cotan formulation.

# 2. Laplace-Beltrami eigenfunctions

Let  $\Delta$  be the Laplace-Beltrami (LB) operator in a reasonably smooth manifold  $\mathcal{M}$  in  $\mathbb{R}^n$ . The LB-operator associated with the Riemannian metric  $g = (g_{ij})$  is then given by [29, 33]

$$\Delta = \frac{1}{|g|^{1/2}} \sum_{i,j} \frac{\partial}{\partial x_i} \Big( |g|^{1/2} g^{ij} \frac{\partial}{\partial x_j} \Big).$$
 (2.1)

Solving the eigenvalue equation

$$\Delta \psi_j(p) = \lambda \psi_j(p), \ p \in \mathcal{M}, \tag{2.2}$$

we obtain ordered eigenvalues

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

and corresponding eigenfunctions  $\psi_0, \psi_1, \psi_2, \ldots$  The first eigenvalue and eigenfunction are trivially given as  $\lambda_0 = 0$  and  $\psi_0 = 1/\sqrt{\mu(\mathcal{M})}$ , where  $\mu(\mathcal{M})$  is the volume of  $\mathcal{M}$ . It is possible to have multiple eigenfunctions corresponding to the same eigenvalue. The multiplicity usually happens when  $\mathcal{M}$  has symmetry. However, if there is no symmetry, all the eigenvalues

#### M. K. Chung & Y. Wang



Fig. 2. Eigenfunctions of the Laplacian in L-shaped domain.

are unique. The eigenfunctions  $\psi_j$  form an orthonormal basis in  $L^2(\mathcal{M})$ [30, 38].

Other than  $\psi_0 = 1/\sqrt{\mu(M)}$ , there is no known close form expressions for eigenfunctions. For most shape modeling applications in medical imaging and computer vision, the underlying domains can be discretized using triangular or tetrahedral meshes; we can discretize the Laplace-Beltrami operator using the cotan formulation as the generalized eigenvalue problem<sup>a</sup> [18, 35, 40]:

$$\mathbf{C}\psi = \lambda \mathbf{A}\psi,\tag{2.3}$$

where **C** is the stiffness matrix, **A** is the mass matrix, and  $\psi$  is the unknown eigenfunction evaluated at mesh vertices. The details on the matrices **C** and **A** can be found in [11, 18, 35]. For tetrahedral meshes in  $\mathbb{R}^3$ , a similar cotan discretization is available [20, 47, 48]. Figure 1 displays few LB-eigenfunctions on the triangle meshes of human amygdala. Figure 2 displays the LB-eigenfunctions on the L-shaped domain.

## 2.1. Graph Laplacian

In computer vision and other sciences, the discrete version of Laplace-Beltrami operator is often used on graph data structures. Even if we do

<sup>&</sup>lt;sup>a</sup>MATLAB code is available at http://brainimaging.waisman.wisc.edu/~chung/lb.

not have graphs, by subdividing regions into a collection of connected polygons, we can obtain graph Laplacian. Almost all the result of continuous counterpart is carried over to the discrete version. Here we explain the main difference in the graph Laplacian.

Let  $G = \{V, E\}$  be a graph with vertex set V and edge set E. We will simply index the node set as  $V = \{1, 2, ..., p\}$ . If two nodes i and j form an edge, we denote it as  $i \sim j$ . Let  $W = (w_{ij})$  be the edge weight. The adjacency matrix of G is often used as the edge weight. Various forms of graph Laplacian have been proposed [10] but the most often used standard form  $L = (l_{ij})$  is given by

$$l_{ij} = \begin{cases} -w_{ij}, & i \sim j\\ \sum_{i \neq j} w_{ij}, & i = j\\ 0, & \text{otherwise.} \end{cases}$$
(2.4)

Even cotan discretization of LB-operator can be written in the form (2.4) [11, 18]. The graph Laplacian L can then be written in a matrix form

$$L = D - W,$$

where  $D = (d_{ij})$  is the diagonal matrix with  $d_{ii} = \sum_{j=1}^{n} w_{ij}$ . For this chapter, we will simply use the adjacency matrix so that the edge weights  $w_{ij}$  are either 0 or 1.

Unlike the continuous Laplace-Beltrami operators that may have possibly infinite number of eigenfunctions, we have up to p number of eigenvectors

ų

$$\psi_1,\psi_2,\ldots,\psi_p$$

satisfying

$$L\psi_j = \lambda_j \psi_j \tag{2.5}$$

with

$$0 = \lambda_1 < \lambda_2 \le \dots \le \lambda_p.$$

The eigenvectors are orthonormal, i.e.,  $\psi_i^{\mathsf{T}}\psi_j = \delta_{ij}$ , the Kroneker's delta. The first eigenvector is trivially given as  $\psi_1 = \mathbf{1}/\sqrt{p}$  with  $\mathbf{1} = (1, 1, \dots, 1)^{\mathsf{T}}$ .

All other higher order eigenvalues and eigenvectors are unknown analytically and have to be computed numerically. Using the eigenvalues and eigenvectors, the graph Laplacian can be decomposed spectrally. From (2.5),

$$L\Psi = \Psi\Lambda, \tag{2.6}$$

where  $\Psi = [\psi_1, \ldots, \psi_p]$  and  $\Lambda$  is the diagonal matrix with entries  $\lambda_1, \ldots, \lambda_p$ . Since  $\Psi$  is an orthogonal matrix,

$$\Psi \Psi^{\mathsf{T}} = \Psi^{\mathsf{T}} \Psi = \sum_{j=1}^{p} \psi_{j} \psi_{j}^{\mathsf{T}} = I_{p},$$

the identify matrix of size p. Then (2.6) is written as

$$L = \Psi \Lambda \Psi^{\mathsf{T}} = \sum_{j=1}^{p} \lambda_j \psi_j \psi_j^{\mathsf{T}}.$$

This is the restatement of the singular value decomposition (SVD) for Laplacian.

### 2.2. Laplacian in positive definite symmetric matrices

Recently in relation to brain network analysis and in diffusion tensor imaging, positive definite symmetric (PDS) matrices have become fundamental object of interest. PDS matrices would be considered as an irregular domain since the usual Euclidean geometry does not apply. Let  $\mathcal{P}_m \subset \mathbb{R}^{m(m+1)/2}$ be the space of positive definite symmetric matrices of size  $m \times m$ .

For  $Y = (y_{ij}) \in \mathcal{P}_m$ , let  $dY = (dy_{ij})$ . Following [32], we will put the following metric on  $\mathcal{P}_m$ :

$$(ds)^2 = \operatorname{tr}\left((Y^{-1}dY)^2\right).$$

Vectorize n = m(m+1)/2 unique entries of Y as  $(x_1, x_2, \ldots, x_n)'$  and write  $ds^2$  in the standard quadratic form as

$$(ds)^2 = \sum g_{ij} dx_i dx_j.$$

For  $\mathcal{P}_m$ , this can be more compactly written as follows. Define the matrix of differential operators  $\partial$  as

$$\partial_Y = \left(\frac{1}{2}(1+\delta_{ij})\frac{\partial}{\partial y_{ij}}\right),$$

where  $\delta_{ij}$  is Kronecker's delta. With this operator, the LB-operator  $\Delta$  in the local coordinates  $y_{ij}$  is given by [24, 36]

$$\Delta = \operatorname{tr}(Y\partial_Y)^2. \tag{2.7}$$

#### Heat Kernel Smoothing in Irregular Domains



Fig. 3. Heat kernel in the L-shaped domain for two different bandwidths  $\sigma = 0.01, 0.1$ . We have used degree 70 expansions but the shape is almost identical if we use higher degree expansions.

Note that the Laplacian in the coordinates of the eigenvalues of y is more complicated from [28]. The eigenfunction of the Laplacian (2.7) is difficult to compute in practice and involves Zonal spherical functions [36, 37].

**Example 2.1:** Consider  $\mathcal{P}_1 = \mathbb{R}^+$ , the positive real line. Note that the Laplacian is parameterization invariant. Let  $y = e^x$  to be the parameterization of  $\mathcal{P}_1$ . It maps  $\mathbb{R}$  to  $\mathbb{R}+$ . Then dy = ydx and with respect to the original coordinates y, we obtain (2.8).

$$\Delta = \left(\frac{d}{dx}\right)^2 = \left(y\frac{d}{dy}\right)^2 = y\frac{d}{dy} + y^2\frac{d^2}{dy^2}.$$
(2.8)

Note Laplacian (2.8) in  $\mathcal{P}_1$  differs from the usual Laplacian  $\frac{d^2}{dy^2}$  for the whole real line. This additional algebraic complexity of the Laplacian makes the computation of eigenfunctions of even 1D case (2.8) complicated. In fact, we need to solve

$$y\frac{d}{dy}\psi_j(y) + y^2\frac{d^2}{dy^2}\psi_j(y) = \lambda_j\psi_j(y).$$
(2.9)

In practice, it might be much easier to simply discretize the differential equation (2.9) and solve using the finite element method.

## 3. Heat kernel

The heat kernel  $K_{\sigma}(p,q)$  is defined as

$$K_{\sigma}(p,q) = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} \psi_j(p) \psi_j(q), \qquad (3.1)$$

where  $\sigma$  is the bandwidth of the kernel. The detailed mathematical exposition of heat kernel is given in [5] and [38]. Note that the heat kernel is the fundamental solution of an isotropic heat diffusion.

Symmetric kernel G(p,q) defined on  $\mathcal{M}$  is positive definite if

$$\sum_{i,j=1}^n G(p_i,q_j)c_ic_j > 0$$

for all choices of  $p_i, q_j \in \mathcal{M}$  and nonzero  $c_i, c_j \in \mathbb{R}$ . Generalizing the definition, kernel G(p,q) is integrally positive definite on  $\mathcal{M}$  if

$$\int_{\mathcal{M}} G(p,q) f(p) f(q) \ d\mu(p) d\mu(q) > 0$$

for any  $f \in L_1(\mathcal{M})$ , the space of integrable functions. For a continuous kernel, these two definitions can be shown to be equivalent.

The heat kernel is a probability distribution, i.e.,

$$\int_{\mathcal{M}} K_{\sigma}(p,q) \ d\mu(p) = \int_{\mathcal{M}} K_{\sigma}(p,q) \ d\mu(q) = 1.$$

Thus, the discretized kernel matrix can be viewed as doubly-stochastic.

Figure 3 shows the heat kernel for bandwidth 0.01 and 1 for L-shaped domain. The kernel follows the shape of the irregular domain.

# 3.1. Heat kernel on spheres

On a two-sphere, the heat kernel is analytically given in terms of the spherical harmonics  $Y_{lm}$  [14]:

$$K_{\sigma}(p,q) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} e^{-l(l+1)\sigma} Y_{lm}(p) Y_{lm}(q).$$
(3.2)

On a three-sphere, the heat kernel is analytically given in terms of the hyperspherical harmonics  $Z_{nlm}$  [21, 27, 26, 25]:

$$K_{\sigma}(p,q) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} e^{-l(l+2)\sigma} Z_{lmn}(p) Z_{lmn}(q).$$

The hyperspherical harmonics  $Z_{lmn}(p)$  with  $p = (\beta, \theta, \phi)$  are given by

$$Z_{lmn}(\beta,\theta,\phi) = 2^{l+1/2} \sqrt{\frac{(n+1)\Gamma(n-l+1)}{\pi\Gamma(n+l+2)}} \Gamma(l+1) \sin^{l} \beta \ G_{n-l}^{l+1}(\cos\beta) Y_{lm}(\theta,\phi),$$

where  $(\beta \in [0, \pi], \theta \in [0, \pi], \phi \in [0, 2\pi])$ ,  $G_{n-1}^{l+1}$  are the Gegenbauer (ultraspherical) polynomials, and  $Y_{lm}$  are the 3D spherical harmonics. The Gegenbauer polynomials can be expressed in terms of the Gaussian (ordinary) hypergeometric function:

$$G_{\alpha}^{\lambda}(x) = \frac{\Gamma(\alpha + 2\lambda)}{\alpha!\Gamma(2\lambda)} {}_{2}F_{1}(-\alpha, \alpha + 2\lambda; \lambda + \frac{1}{2}; \frac{1}{2}(1-x)).$$

The hyperpsherical harmonics form an orthonormal basis on the hypersphere:

$$\int_0^{2\pi} \int_0^{\pi} \int_0^{\pi} Z_{lmm}(\beta,\theta,\phi) Z_{l'm'n'}(\beta,\theta,\phi) \sin^2\beta \sin\theta d\beta d\theta d\phi = \delta_{nn'} \delta_{ll'} \delta_{mm'}.$$

## 3.2. Heat kernel on graphs

On a graph, the discrete heat kernel  $K_{\sigma}$  has a finite expansion and the simplicity in the algebraic representation makes it easier to manipulate.  $K_{\sigma}$  is a positive definite symmetric matrix of size  $p \times p$  given by

$$K_{\sigma} = \sum_{j=1}^{p} e^{-\lambda_{j}\sigma} \psi_{j} \psi_{j}^{\mathsf{T}}, \qquad (3.3)$$

where  $\sigma$  is called the bandwidth of the kernel. Alternately, we can write (3.3) as

$$K_{\sigma} = \Psi e^{-\sigma \Lambda} \Psi^{\mathsf{T}},$$

where  $e^{-\sigma\Lambda}$  is the matrix logarithm of  $\Lambda$ . To see positive definiteness of the kernel, for any nonzero  $x \in \mathbb{R}^p$ , note

$$x^{\mathsf{T}} K_{\sigma} x = \sum_{j=1}^{p} e^{-\lambda_{j}\sigma} x^{\mathsf{T}} \psi_{j} \psi_{j}^{\mathsf{T}} x$$
$$= \sum_{j=1}^{p} e^{-\lambda_{j}\sigma} (\psi_{j}^{\mathsf{T}} x)^{2} > 0.$$

When  $\sigma = 0$ ,  $K_0 = I_p$ , identity matrix. When  $\sigma = \infty$ , by interchanging the sum and the limit, we obtain

$$K_{\infty} = \psi_1 \psi_1^{\mathsf{T}} = \mathbf{1} \mathbf{1}^{\mathsf{T}} / p.$$

### M. K. Chung & Y. Wang

 $K_{\infty}$  is a degenerate case and the kernel is no longer positive definite. Other than these specific cases, the heat kernel is not analytically known in arbitrary graphs.

Heat kernel is doubly-stochastic [10] so that

$$K_{\sigma}\mathbf{1} = \mathbf{1}, \ \mathbf{1}^{\mathsf{T}}K_{\sigma} = \mathbf{1}^{\mathsf{T}}.$$

Thus,  $K_{\sigma}$  is a probability distribution along columns or rows.

Just like the continuous counterpart, the discrete heat kernel is also multiscale and has the scale-space property. Note

$$K_{\sigma}^{2} = \sum_{i,j=1}^{p} e^{-(\lambda_{i}+\lambda_{j})\sigma} \psi_{i}\psi_{i}^{\mathsf{T}}\psi_{j}\psi_{j}^{\mathsf{T}}$$
$$= \sum_{j=1}^{p} e^{-2\lambda_{j}\sigma} \psi_{j}\psi_{j}^{\mathsf{T}} = K_{2\sigma}.$$

We used the orthonormality of eigenvectors. Subsequently, we have

 $K_{\sigma}^{n} = K_{n\sigma}$ 

for any integer  $n \ge 0$ .

## 4. Heat kernel smoothing

The concept of heat kernel smoothing was introduced in [17, 16] in the context of smoothing human cortical surface data. The original formulation used the tangent space approximation. The spectral version using LB-eigenfunction was later developed [40, 15]. *Heat kernel smoothing* of functional measurement f(p) is then defined as

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

where

$$f_j = \langle f, \psi_j \rangle = \int_{\mathcal{M}} f(p) \psi_j(p) \ d\mu(p)$$

are Fourier coefficients [17]. It is well known that heat kernel smoothing is the unique solution of an isotropic heat diffusion [38].

**Theorem 4.1:** For an arbitrary self-adjoint differential operator  $\Delta$  and

 $f \in L^2(\mathcal{M})$ , the unique solution of the Cauchy problem

$$\frac{\partial g(p,\sigma)}{\partial \sigma} + \Delta g(p,\sigma) = 0, g(p,\sigma=0) = f(p) \tag{4.1}$$

is given by

$$g(p,\sigma) = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} f_j \psi_j(p).$$
(4.2)

**Proof:** The statement is first given in [13] with heuristic proof. Here we provide a more rigorous proof. We first prove  $g(p, \sigma) \in L^2(\mathcal{M})$ . Since  $g(p, \sigma)$  is the solution to (4.1), multiplying (4.1) with  $g(p, \sigma)$  and integrating on  $[0, T] \times \mathcal{M}$ , we obtain

$$\int_0^T \int_{\mathcal{M}} \frac{\partial g(p,\sigma)}{\partial t} g(p,\sigma) \, d\mu(p) \, d\sigma + \int_0^T \int_{\mathcal{M}} \Delta g(p,\sigma) g(p,\sigma) \, d\mu(p) \, d\sigma = 0,$$

where T > 0 is total diffusion time. Hence, it holds that

$$\frac{1}{2} \int_{\mathcal{M}} \left( g(p,T)^2 - g(p,0)^2 \right) d\mu(p) + \int_0^T \langle \Delta g, g \rangle \, d\sigma = 0.$$

Since  $\Delta$  is self-adjoint,

$$\left< \Delta g, g \right> \geq 0$$

Thus,

$$\begin{split} \|g(\cdot,T)\|_{2}^{2} &= \int_{\mathcal{M}} g(p,T)^{2} \ d\mu(p) \\ &= \int_{\mathcal{M}} g(p,0)^{2} \ d\mu(p) - 2 \int_{0}^{T} \langle \Delta g,g \rangle \ d\sigma \\ &\leq \int_{\mathcal{M}} g(p,0)^{2} \ d\mu(p) = \int_{\mathcal{M}} f(p)^{2} \ d\mu(p) = \|f\|_{2}^{2} < \infty, \end{split}$$

where  $\|\cdot\|_2$  is the  $L^2$ -norm.

Since eigenfunctions  $\psi_j$  form an orthonormal basis in  $\mathcal{M}$ , for each fixed  $\sigma$ ,  $g(p, \sigma)$  can be *uniquely* written as a Fourier series

$$g(p,\sigma) = \sum_{j=0}^{\infty} c_j(\sigma)\psi_j(p).$$
(4.3)

Then

$$\Delta g(p,\sigma) = \sum_{j=0}^{\infty} c_j(\sigma) \lambda_j \psi_j(p).$$
(4.4)

## M. K. Chung & Y. Wang

Substituting (4.3) and (4.4) into (4.1), we obtain

$$\frac{\partial c_j(\sigma)}{\partial \sigma} + \lambda_j c_j(\sigma \Delta a) = 0 \tag{4.5}$$

for all j. The solution of equation (4.5) is given by  $c_j(\sigma) = b_j e^{-\lambda_j \sigma}$ . So we have a 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  $b_i$  must be given by the *unique* Fourier coefficients, i.e.,

$$b_j = \langle f, \psi_j \rangle = f_j.$$

Since heat kernel smoothing is the solution of diffusion equation, it satisfies the scale-space property.

**Theorem 4.2:** Denote the k-fold iterated kernel as

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

Then we have

$$K_{k\sigma} * f = K_{\sigma}^{(k)} * f. \tag{4.6}$$

**Proof:**  $K_{\sigma}^{(2)} * f$  is equivalent to the diffusion of signal f after time  $2\sigma$ . Hence we have

$$K_{\sigma}^{(2)}f = K_{2\sigma} * f.$$

Arguing inductively we see that the general statement holds.

Note an alternate proof can be obtained by noting that  $K_{\sigma}^{(k)}$  is the density of the sum of k independent and identically distributed random variables in  $\mathcal{M}$ . Heat kernel with a large bandwidth is equivalent to the multiple applications of heat kernel smoothing with a smaller bandwidth. The property (4.6) was used to approximate heat kernel smoothing with

#### Heat Kernel Smoothing in Irregular Domains



Fig. 4. Top left to right: 3D lung vessel tree. Gaussian noise is added to one of the coordinates. Heat kernel smoothing with bandwidth 0.01, 0.1, 1 and 10000.

multiple applications of Gaussian kernel smoothing with small bandwidth [17]. If we change the scale to  $2\sigma = \tau^2$ , (4.6) takes a slightly different form:

$$K_{\tau}^{(k)} * f = K_{\sqrt{k\tau}} * f.$$

## 4.1. Asymptotics

As  $\sigma \to 0$ ,  $K_{\sigma}(p,q)$  becomes the Dirac-delta function  $\delta(p-q)$  so the heat kernel smoothing becomes unbiased as  $\sigma \to 0$ , i.e.,

$$\lim_{\sigma \to 0} K_{\sigma} * f(p) = f(p).$$

**Theorem 4.3:** For  $f \in L^2(\mathcal{M})$  with  $\mu(\mathcal{M}) \leq \infty$ , the heat kernel smoothing converges to the mean signal over  $\mathcal{M}$  pointwisely

$$\lim_{\sigma \to \infty} K_{\sigma} * f(p) = \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} f(p) \ d\mu(p),$$

for all  $p \in \mathcal{M}$ .

**Proof:** The statement is given in [17, 38]. Since  $K_{\sigma} * f$  is bounded, we can interchange the limit and summation.

$$\lim_{\sigma \to \infty} K_{\sigma} * f = \lim_{\sigma \to \infty} \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} f_j \psi_j(p)$$
$$= \sum_{j=0}^{\infty} \lim_{\sigma \to \infty} e^{-\lambda_j \sigma} f_j \psi_j(p)$$
$$= f_0 \psi_0(p)$$
$$= \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} f \ d\mu(p).$$

Theorem 4.3 can be used to identify the number of disconnected structures in very complex structures. Figure 4 shows a part of lung vessel tree obtained from computed tomography (CT) [19, 50]. Gaussian noise is added to one of the coordinates. 3D volumetric heat kernel smoothing with bandwidth 0.01, 0.1, 1 and 10000 is performed on voxels. At the bandwidth 10000, heat kernel smoothing is almost reaching the steady state. Differently colored vessel tree shows they are disconnected structures. There are total 7 disconnected structures.

Since  $\psi_0 = 1/\sqrt{\mu(\mathcal{M})}$ , we have

$$K_{\sigma} * f(p) = \frac{\int_{\mathcal{M}} f(p) \, d\mu(p)}{\mu} (\mathcal{M}) + f_1 e^{-\lambda_1 \sigma} \psi_1(p) + R(\sigma, p), \qquad (4.7)$$

where the first term is the average signal,  $f_1$  is a constant and the remaining term R goes to 0 faster than  $e^{-\lambda_1 \sigma}$  as  $\sigma \to \infty$  [2]. Due to expansion (4.7), the behavior of heat kernel smoothing is basically governed by the second eigenfunction  $\psi_1$  for large bandwidth.

#### 4.2. Inequalities

We are interested in bounding heat kernel smoothing  $K_{\sigma} * f(p)$ . We present few useful inequalities involving heat kernel smoothing.

**Theorem 4.4:** Conservation of signal:

$$\int_{\mathcal{M}} K_{\sigma} * f(p) \ d\mu(p) = \int_{\mathcal{M}} f(p) \ d\mu(p).$$



Fig. 5. Heat kernel smoothing on the surface coordinates of hippocampus mesh with different bandwidth.

**Proof:** This is due to the fact that  $K_{\sigma}$  is a probability distribution in  $\mathcal{M}$ , i.e.,

$$\int_{\mathcal{M}} K_{\sigma}(p,q) \ d\mu(q) = 1$$

for all  $p \in \mathcal{M}$ . Then

August 29, 2019

$$\int_{\mathcal{M}} K_{\sigma} * f(p) \ d\mu(p) = \int_{\mathcal{M}} f(p) \int_{\mathcal{M}} K_{\sigma}(p,q) \ d\mu(q) \ d\mu(p)$$
$$= \int_{\mathcal{M}} f(p) \ d\mu(p).$$

195

**Theorem 4.5:** For  $f \in L^1(\mathcal{M})$ ,

$$||K_{\sigma} * f||_1 \le ||f||_1.$$

**Proof:** For  $f \ge 0$ , from Theorem 4.4, trivially we have equality

$$||K_{\sigma} * f||_1 = ||f||_1.$$

If  $f \leq 0$ , then |f| = -f and we have the same result. If the sign of f is not constant, consider decomposition  $f = f^+ + f^-$ , where

$$f^{+}(p) = \begin{cases} f(p), f(p) > 0, \\ 0, f(p) \le 0. \end{cases} \quad f^{-}(p) = \begin{cases} f(p), f(p) < 0, \\ 0, f(p) \ge 0. \end{cases}$$

We can write |f| as

$$|f| = |f^+| + |f^-|.$$

Then we have

$$\|K_{\sigma} * f\|_{1} = \int_{\mathcal{M}} \left| K_{\sigma} * f^{+}(p) + K_{\sigma} * f^{-}(p) \right| d\mu(p)$$

$$\leq \int_{\mathcal{M}} \left| K_{\sigma} * f^{+}(p) \right| + \left| K_{\sigma} * f^{-}(p) \right| d\mu(p)$$

$$= \|f^{+}\|_{1} + \|f^{-}\|_{1}$$

$$= \|f\|_{1}.$$

Only when the sign of f does not change, Theorem 4.5 is equality. Consider the following counter example. Consider  $\mathcal{M}_1, \mathcal{M}_2 \subseteq \mathcal{M}, \mathcal{M}_1 \cap \mathcal{M}_2 = \emptyset, \mathcal{M}_1 \cup \mathcal{M}_2 = \mathcal{M}$ , such that

$$\int_{\mathcal{M}_1} K_{\sigma}(p,q) \ d\mu(q) = 0.5, \quad \int_{\mathcal{M}_2} K_{\sigma}(p,q) \ d\mu(q) = 0.5.$$

Set f(p) as

$$f(p) = \begin{cases} 1, & p \in \mathcal{M}_1, \\ -1, & p \in \mathcal{M}_2. \end{cases}$$

Then

$$||K_{\sigma} * f||_{1} = || \int_{\mathcal{M}_{1}} K_{\sigma}(p,q)(1) \ d\mu(q) + \int_{\mathcal{M}_{2}} K_{\sigma}(p,q)(-1) \ d\mu(q) ||_{1}$$
$$= ||0.5 + (-0.5)||_{1} = 0.$$

While

$$||f||_1 = \int_{\mathcal{M}} 1 \ d\mu(p) = \mu(\mathcal{M}).$$

We prove similar result for other  $L^n$ -norms.

**Theorem 4.6:** For  $f \in L^n(\mathcal{M})$ , we have norm contraction

$$\|K_{\sigma_1} * f\|_n \le \|K_{\sigma_2} * f\|_n \le \cdots$$

for  $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$  and  $n \geq 2$ .

**Proof:** Consider  $g \in L^n(\mathcal{M})$ . Based on Jensen's inequality and the fact that  $K_{\sigma}$  is a probability distribution, we have

$$\|K_{\sigma} * g\|_{n}^{n} = \int_{\mathcal{M}} \left| \int_{\mathcal{M}} K_{\sigma}(p,q)g(q) \, d\mu(q) \right|^{n} \, d\mu(p)$$
  
$$\leq \int_{\mathcal{M}} \int_{\mathcal{M}} K_{\sigma}(p,q)|g(q)|^{n} \, d\mu(q) \, d\mu(p)$$
  
$$= \|K_{\sigma} * (|g|^{n})\|_{1} = \|g\|_{n}^{n}.$$
(4.9)

Equation (4.9) is due to Theorem 4.5. Now let  $\sigma = \sigma_1$  and  $g = K_{\sigma_1 - \sigma_2} * f$ . Then we have

$$||K_{\sigma_1} * f||_n^n = ||K_{\sigma_1 - \sigma_2} * (K_{\sigma_2} * f)||_n^n \le ||K_{\sigma_2} * f||_n^n.$$

For  $L^2$ -norm, Theorem 4.6 is the consequence of the Hilbert space isomorphism. Note

$$||K_{\sigma} * f||_{2} = ||\sum_{j} e^{-\lambda_{j}\sigma} f_{j}\psi_{j}||_{2} = \sum_{j=0}^{\infty} e^{-2\lambda_{j}\sigma} f_{j}^{2}.$$

Since  $\sum_{j=0}^{\infty} e^{-2\lambda j\sigma} f_j^2 \leq \sum_{j=0}^{\infty} f_j^2$ , we have  $||K_{\sigma} * f||_2 \leq ||f||_2$ . Theorem 4.6 can be used to bound  $K_{\sigma} * f(p)$  uniformly. From Hölder's inequality and from Theorem 4.6, for each fixed p,

$$\begin{aligned} \left| K_{\sigma} * f(p) \right| &\leq \int_{\mathcal{M}} K_{\sigma}(p,q) \left| f(q) \right| d\mu(q) \\ &\leq \mu(\mathcal{M})^{1/2} \| K_{\sigma} * f \|_{2} \\ &\leq \mu(\mathcal{M})^{1/2} \| f \|_{2}. \end{aligned}$$
(4.10)

Note we also have  $|K_{\sigma} * f(p)| \leq \sup_{p \in \mathcal{M}} |f(p)|.$ 

**Theorem 4.7:** For  $f \in L^{\infty}(\mathcal{M})$ , we have norm contraction

$$||K_{\sigma_1} * f||_{\infty} \le ||K_{\sigma_2} * f||_{\infty} \le \cdots$$

for  $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ .

**Proof:** Consider  $g \in L^{\infty}(\mathcal{M})$ . We have

$$|K_{\sigma} * g(p)| \leq \int_{\mathcal{M}} K_{\sigma}(p,q) |g(q)| \ d\mu(q)$$
$$\leq ||g||_{\infty} \int_{\mathcal{M}} K_{\sigma}(p,q) \ d\mu(q)$$
$$= ||g||_{\infty}$$

for any  $p \in \mathcal{M}$ . Hence,

$$||K_{\sigma} * g||_{\infty} = \sup_{p \in \mathcal{M}} |K_{\sigma} * g(p)| \le ||g||_{\infty}.$$

Now let  $\sigma = \sigma_1$  and  $g = K_{\sigma_1 - \sigma_2} * f$ . Then we have

 $||K_{\sigma_1} * f||_{\infty} = ||K_{\sigma_1 - \sigma_2} * (K_{\sigma_2} * f)||_{\infty} \le ||K_{\sigma_2} * f||_{\infty}.$ 

#### M. K. Chung & Y. Wang

The result of smoothing is often used in quantifying a collection of shapes and functions in medical imaging [16, 15]. Thus, we investigate the distance between two different heat kernel smoothings and how it changes from before smoothing. We define  $L^n$ -distance between functions  $f, g \in$  $L^n(\mathcal{M})$  as that

$$d_n(f,g) = ||f-g||_n.$$

Then we can show that the distance between functions decreases after heat kernel smoothing.

**Theorem 4.8:** Heat kernel smoothing is a contraction map in  $L^n(\mathcal{M})$  in a sense that for any  $f, g \in L^n(\mathcal{M})$ ,

$$d_n(K_{\sigma_1} * f, K_{\sigma_1} * g) \le d_n(K_{\sigma_2} * f, K_{\sigma_2} * g)$$

for  $\sigma_1 \geq \sigma_2 \geq 0$  and  $n \geq 1$ .

**Proof:** It follows from Theorems 4.2 (scale-space property) that

$$K_{\sigma_1} * f = K_{\sigma_1 - \sigma_2} * K_{\sigma_2} * f.$$

From Theorem 4.6,

$$||K_{\sigma_1} * f||_n = ||K_{\sigma_1 - \sigma_2} * (K_{\sigma_2} * f)||_n \le ||K_{\sigma_2} * f||_n$$

for  $\sigma_1 \geq \sigma_2 \geq 0$ .

As a special case of Theorem 4.8, we have

$$d_n(K_{\sigma} * f, K_{\sigma} * g) \le d_n(f, g).$$

Theorem 4.8 also holds true for  $L^{\infty}$ -norm as well.

## 4.3. Discrete heat kernel smoothing on graphs

Discrete heat kernel smoothing of measurement vector  $f = (f_1, f_2, \dots, f_p)^{\mathsf{T}}$ on a graph is defined similarly as

$$K_{\sigma} * f = K_{\sigma}f = \sum_{j=0}^{p} e^{-\lambda_{j}\sigma} \tilde{f}_{j}\psi_{j}.$$
(4.11)

This is the discrete analogue of heat kernel smoothing first defined in [17]. In discrete setting, the convolution \* is simply a matrix multiplication. Then

$$K_0 * f = f$$

#### Heat Kernel Smoothing in Irregular Domains

and

$$K_{\infty} * f = f\mathbf{1},$$

where  $\bar{f} = \sum_{j=1}^{p} f_j/p$  is the mean of signal f over every node. When the bandwidth is zero, we are not smoothing data. As the bandwidth increases, the smoothed signal converges to the sample mean of all values. Then we have similar results for the discrete version as well. Here we only show the contraction mapping property simply to illustrate the differences.

**Theorem 4.9:** Heat kernel smoothing is a contraction mapping with respect to the  $l^n$ -norm for vectors, i.e.,

$$||K_{\sigma} * f||_{n}^{n} \le ||f||_{n}^{n}.$$

**Proof:** Let kernel matrix  $K_{\sigma} = (k_{ij})$ . Then we have inequality

$$||K_{\sigma} * f||_{n}^{n} = \sum_{i=1}^{p} \sum_{j=1}^{p} |k_{ij}f_{j}|^{n} \le \sum_{j=1}^{p} |f_{j}|^{n}.$$

We used Jensen's inequality and doubly-stochastic property of the heat kernel.  $\hfill \Box$ 

A similar result can be obtained for  $l^{\infty}$ -norm.

## 5. Statistical properties of heat kernel smoothing

Often observed noisy data on graphs is smoothed to increase the signalto-noise ratio (SNR) and increases the statistical sensitivity [15]. We are interested in knowing how heat kernel smoothing will have effects on the statistical properties on the data. In practice, functional data Y(p) is modeled as a random field:

$$Y(p) = f(p) + e(p),$$
 (5.1)

where unknown deterministic signal  $f \in L^2(\mathcal{M})$  and  $\epsilon$  is a zero-mean random field with some covariance function  $R_e(p,q)$ , i.e.,

$$R_e(p,q) = \mathbb{E}[e(p)e(q)].$$

We will further assume constant variance field, i.e.,

$$R_e(p,p) = \mathbb{V}e(p) = \mathbb{E}e^2(p) = \text{const}$$

for all  $p \in \mathcal{M}$ .

## M. K. Chung & Y. Wang

The covariance functions are often unimodal and isotropic in  $\mathcal{M}$ . A function is isotropic a manifold in the following sense. Consider line segment  $C \subset \mathcal{M}$  connecting p and q and parameterized by  $\gamma_c(t)$  with  $\gamma_c(0) = p$  and  $\gamma_c(1) = q$ . In the Cartesian coordinates,  $\gamma_c(t) = (\gamma_c^1(t), \ldots, \gamma_c^n(t)) \in \mathbb{R}^n$ . The length of C is given by

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

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

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

It is usually given as the solution of the Euler equation and numerical techniques are available for polygonal surfaces [49]. Suppose the covariance function of e is of unimodal isotropic function of the form  $R_e(p,q) = \rho(d(p,q))$ where d(p,q) is the geodesic distance between p and q and  $\rho$  is some nonincreasing function. This is an often encountered covariance function shape in applications. Note d(p,p) = 0 and  $R_e(p,p) = \rho(0)$ . Noise e(p) can be further modeled as Gaussian white noise, i.e., Brownian motion or the generalized derivatives of Wiener process, whose covariance function is Dirac-delta, i.e.,

$$R_e(p,q) = \delta(p-q).$$

Often observed functional data Y(p) is smoothed with heat kernel  $K_{\sigma}$  to increase the signal-to-noise ratio (SNR) and increases the statistical sensitivity [15]. Once heat kernel smoothing is applied to (5.1), we have

$$K_{\sigma} * Y(p) = K_{\sigma} * f(p) + K_{\sigma} * e(p).$$
(5.2)

For Gaussian white noise e, its covariance function of  $K_{\sigma} * e$  is given by

$$R_{K_{\sigma}*e}(p,q) = \int_{\mathcal{M}} K_{\sigma}(p,r) K_{\sigma}(q,r) \ d\mu(r).$$

The variance at p is then

$$\mathbb{V}[K_{\sigma} * e(p)] = R_{K_{\sigma} * e}(p, p) = \int_{\mathcal{M}} K_{\sigma}^{2}(p, r) \ d\mu(r).$$

The variance of data will be often reduced after heat kernel smoothing in the following sense [17, 16]. This is formulated rigorously as follows.

We are interested in determining how the statistical properties of signal change between (5.1) and (5.10). This is needed to study the behavior of collection of heat kernel smoothed functions statistically. We can show that the variance of smoothed noise is smaller than the variance of noise.

## Theorem 5.1:

$$\mathbb{V}[K_{\sigma} * Y(p)] \le \mathbb{V}Y(p)$$

for all  $p \in \mathcal{M}$ .

**Proof:** Note that

$$\mathbb{V}e(p) = \mathbb{V}Y(p)$$
$$\mathbb{V}[K_{\sigma} * Y(p)] = \mathbb{V}[K_{\sigma} * e(p)].$$

Since  $\mathbb{E}(K_{\sigma} * e(p)) = 0$ ,

$$\mathbb{V}[K_{\sigma} * e(p)] = \mathbb{E}\left[\left(K_{\sigma} * e(p)\right)^{2}\right]$$

It follows from Theorem 4.4 and Jensen's inequality that

$$\mathbb{E}\Big[\int_{\mathcal{M}} K(p,q)e(q) \ d\mu(q)\Big]^2 \leq \mathbb{E}\Big[\int_{\mathcal{M}} K(p,q)e(q)^2 \ d\mu(q)\Big]$$
$$= \mathbb{E}e^2(p) \int_{\mathcal{M}} K(p,q) \ d\mu(q)$$
$$= \mathbb{E}e^2(p).$$

Theorem 5.1 shows heat kernel smoothing reduces the point-wise variability of functional signal. Note the usual *t*-statistic often used in anatomical shape discrimination analysis [16, 15] is inversely proportional to the standard deviation. Since heat kernel smoothing reduces the variability, *t*-statistics will likely increase.

#### 5.1. Heat kernel regression on manifolds

Consider subspace  $\mathcal{H}_k \subset L^2(\mathcal{M})$  spanned by the orthonormal basis  $\{\psi_j\}$ , i.e.,

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

## M. K. Chung & Y. Wang

Then the least squares estimation (LSE) of unknown signal f in  $\mathcal{H}_k$  in model (5.1) is given by the shortest distance from observed signal Y to  $\mathcal{H}_k$ :

$$\widehat{f}(p) = \arg\min_{h \in \mathcal{H}_k} \int_{\mathcal{M}} \left| Y(p) - h(p) \right|^2 d\mu(p) = \sum_{j=0}^k Y_j \psi_j(p), \qquad (5.3)$$

where  $Y_j = \langle Y, \psi_j \rangle$  are the Fourier coefficients. This is the usual Fourier series expansion that tends to suffer the Gibbs phenomenon, i.e., ringing artifact, for compact surfaces [13, 22]. The Gibbs phenomenon can be effectively removed if the Fourier series expansion converges fast enough as the number of basis functions goes to infinity. By weighting the Fourier coefficients exponentially smaller, we can make the representation converges faster; this can be achieved by additionally weighting the squared residuals in equation (5.3) with the heat kernel:

$$\widehat{f}(p) = \arg\min_{h \in \mathcal{H}_k} \int_{\mathcal{M}} \int_{\mathcal{M}} K_{\sigma}(p,q) \left| Y(q) - h(p) \right|^2 d\mu(q) \ d\mu(p).$$
(5.4)

The optimization (5.4) has the following analytic expression:

## Theorem 5.2:

$$\widehat{f}(p) = \arg\min_{h \in \mathcal{H}_k} \int_{\mathcal{M}} \int_{\mathcal{M}} K(p,q) |Y(q) - h(p)|^2 d\mu(q) d\mu(p) = \sum_{j=0}^k \tau_j Y_j \psi_j,$$

where  $Y_j = \langle Y, \psi_j \rangle$  are Fourier coefficients.

**Proof:** Any function  $h \in \mathcal{H}_k$  can be expressed as

$$h(p) = \sum_{j=0}^{k} \beta_j \psi_j(p).$$
(5.5)

Then by plugging (5.5) into the inner integral I(p), it becomes

$$I(p) = \int_{\mathcal{M}} K_{\sigma}(p,q) \left| Y(q) - \sum_{j=0}^{k} \beta_{j} \psi_{j}(p) \right|^{2} d\mu(q).$$

Simplifying the expression, we obtain

$$I(p) = \sum_{j=0}^{k} \sum_{j'=0}^{k} \psi_j(p)\psi_{j'}(p)\beta_j\beta_{j'} - 2K_\sigma * Y(p)\sum_{j=0}^{k} \psi_j(p)\beta_j + K_\sigma * Y^2(p).$$
(5.6)

The kernel can be written as

$$K_{\sigma}(p,q) = \sum_{j'=0}^{\infty} \tau_{j'} \psi_{j'}(p) \psi_{j'}(q)$$
(5.7)

and the convolution is written as

$$K_{\sigma} * Y(p) = \sum_{j'=0}^{\infty} \tau_{j'} Y_{j'} \psi_{j'}(p).$$

Since I is an unconstrained positive semidefinite quadratic program (QP) in  $\beta_j$ , there is no unique global minimizer of I without additional linear constraints. Integrating I further with respect to  $d\mu(p)$ , we collapse (5.6) to a positive definite QP, which yields a unique global minimizer:

$$\int_{\mathcal{M}} I(p) \ d\mu(p) = \sum_{j=0}^{k} \beta_j^2 - 2 \sum_{j=0}^{k} \tau_j Y_j \beta_j + \text{ const.}$$

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

$$\int_{\mathcal{M}} \frac{\partial I}{\partial \beta_j} d\mu(p) = 2\beta_j - 2\tau_j Y_j = 0$$

for all j. Hence  $\sum_{j=0}^{k} \tau_j Y_j \psi_j$  must be the unique minimizer.

Theorem 5.2 generalizes the weighted spherical harmonic (SPHARM) representation on a unit sphere to an arbitrary manifold [14]. Theorem 5.2 implies that the kernel regression can be performed by simply computing the Fourier coefficients  $f_j = \langle f, \psi_j \rangle$  without doing any numerical optimization. The numerically difficult optimization problem is reduced to the problem of computing Fourier coefficients. If the kernel K is a Diracdelta function, the kernel regression simply collapses to the least squares estimation (LSE) which results in the standard Fourier series, i.e.

$$\widehat{f}(p) = \arg\min_{h \in \mathcal{H}_k} \int_{\mathcal{M}} \left| Y(q) - h(q) \right|^2 d\mu(q) = \sum_{j=0}^k f_j \psi_j.$$

It can be also shown that as  $k \to \infty$ , the kernel regression

$$\widehat{f} = \sum_{j=0}^{k} \tau_j Y_j \psi_j$$

203

converges to convolution  $K_{\sigma} * Y$  establishing the connection to the manifoldbased kernel smoothing framework [3, 17]. Hence, asymptotically the proposed kernel regression should inherit many statistical properties of kernel smoothing.

## 5.2. Statistical properties on graphs

Similar results can be obtained for graph data structures. Consider the following model:

$$f = \mu + e$$

where  $\mu$  is an unknown signal and  $\epsilon$  is zero mean noise. Let  $e = (e_1, \ldots, e_p)^{\mathsf{T}}$ . Denote  $\mathbb{E}$  as expectation and  $\mathbb{V}$  as covariance. It is natural to assume that the variability of noises at different nodes j is identical, i.e.,

$$\mathbb{E}e_1^2 = \mathbb{E}e_2^2 = \dots = \mathbb{E}e_p^2.$$
(5.8)

Further, we assume that data at two nodes i and j to have less correlation when the distance between the nodes is large. So covariance matrix  $R_e = \mathbb{V}e = \mathbb{E}(ee^{\mathsf{T}}) = (r_{ij})$  can be given by

$$r_{ij} = \rho(d_{ij}) \tag{5.9}$$

for some decreasing function  $\rho$  and geodesic distance  $d_{ij}$  between nodes i and j. Note  $r_{jj} = \rho(0)$  with the understanding that  $d_{jj} = 0$  for all j. The off diagonal entries of  $R_e$  are smaller than the diagonal. Noise e can be further modeled as the discrete Gaussian white noise whose covariance matrix elements are Kroneker-delta  $\delta_{ij}$  with  $\delta_{ij} = 1$  if i = j and 0 otherwise. Thus,

$$R_e = \mathbb{E}(ee^\mathsf{T}) = I_p$$

the identity matrix of size  $p \times p$ . Since  $\delta_{jj} \ge \delta_{ij}$ , Gaussian white noise is a special case of (5.9). After heat kernel smoothing, we have

$$K_{\sigma} * f = K_{\sigma} * \mu + K_{\sigma} * e. \tag{5.10}$$

For  $R_e = I_p$ , the covariance matrix of smoothed noise is simply given as

$$R_{K_{\sigma}*e} = K_{\sigma} \mathbb{E}(ee^{\mathsf{T}}) K_{\sigma} = K_{\sigma}^2 = K_{2\sigma}.$$

We used the scale-space property of heat kernel. In general, the covariance matrix of smoothed data  $K_{\sigma} * e$  is given by

$$R_{K_{\sigma}*e} = K_{\sigma} \mathbb{E}(ee^{\mathsf{T}}) K_{\sigma} = K_{\sigma} R_e K_{\sigma}$$

Other than these differences, similar analogous results can be obtained.

August 29, 2019

### 5.3. Persistent homology in heat kernel smoothing

In persistent homology, a point cloud is used to build a Rips filtration [1, 7, 9]. Similarly, we can build Rips filtration in a function space. Given a collection of functional measurements in  $L^n(\mathcal{M})$ , heat kernel smoothing induces a Rips filtration in  $L^n(\mathcal{M})$  if we take the functions as a point cloud and build a filtration using  $L^n$ -norm as distance.

**Theorem 5.3:** Let  $A_{\sigma} = \{f \in L^n(\mathcal{M}) : ||K_{\sigma} * f||_n \leq h\}$ . Then  $A_{\sigma}$  induces filtration

$$A_{\sigma_1} \subset A_{\sigma_2} \subset \cdots$$

for any  $\sigma_1 \ge \sigma_2 \ge \cdots \ge 0$ ,  $h \ge 0$  and  $n \ge 1$ .

**Proof:** Suppose  $f \in A_{\sigma_1}$ . From Theorem 4.6 (norm contraction),

$$|K_{\sigma_1} * f||_n \le ||K_{\sigma_2} * f||_n \le h.$$

Then  $f \in A_{\sigma_2}$  and the result follows.

A similar result can be obtained for  $L^{\infty}(\mathcal{M})$  space as well. Theorem 5.3 build filtrations on the space of functions. We can also build a filtration directly in manifold  $\mathcal{M}$  as well.

**Theorem 5.4:** Let 
$$B_{\sigma} = \{p \in \mathcal{M} : \mathbb{V}[K_{\sigma} * Y(p)] \leq h\}$$
. Then  $B_{\sigma}$  satisfies  
 $B_{\sigma_1} \subset B_{\sigma_2} \subset \cdots B_0$  (5.11)

if  $\sigma_1 \ge \sigma_2 \ge \cdots \ge 0$  for any  $h \ge 0$  and  $n \ge 1$ .

**Proof:** Let  $p \in B_{\sigma_1}$ . Then from Theorem 5.1,

$$\mathbb{V}[K_{\sigma_1} * Y(p)] = \mathbb{V}[K_{\sigma_1 - \sigma_2} * (K_{\sigma_2} * Y)(p)]$$
$$\leq \mathbb{V}[K_{\sigma_2} * Y(p)] \leq h.$$

Thus,  $p \in B_{\sigma_2}$ .

## 6. Discussion

For irregular domains in images, boundary shapes are often complex. This causes the geometric shape of the boundary to strongly bias smoothing. [43] proposed more natural boundary conditions that reduces the boundary induced bias in smoothing by using the Neumann boundary condition in solving a partial different equation. The heat kernel smoothing method

proposed here is based on the Dirichlet boundary condition although extending it to the Neumann boundary condition is also possible. For closed surfaces with no boundary, there is no need to consider for the boundary condition.

#### Acknowledgment

This work was partially supported by the NIH research grant R01 EB022856. We would like to thank Jim Ramsay of McGill University, Michelle Carey of University College Dublin and Yu-Min Chung of College of Willam and Mary for valuable discussions on heat kernel smoothing in irregular domains. We would like to thank Ruth Sullivan, Michael Johnson and Michael Newton of University of Wisconsin-Madison for useful discussions on modeling vessel trees. Gurong Wu of University of North Carolina-Chapel Hill provided the lung vessel tree data illustrated in this chapter.

## References

- R.J. Adler, O. Bobrowski, M.S. Borman, E. Subag, and S. Weinberger. Persistent homology for random fields and complexes. In *Borrowing strength: theory powering applications-a Festschrift for Lawrence D. Brown*, pages 124–143. Institute of Mathematical Statistics, 2010.
- R. Banuelos and K. Burdzy. On the Hot Spots Conjecture of J. Rauch. Journal of Functional Analysis, 164:1–33, 1999.
- M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In Advances in Neural Information Processing Systems, pages 585–592, 2002.
- M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *The Journal of Machine Learning Research*, 7:2399–2434, 2006.
- 5. N. Berline, E. Getzler, and M. Vergne. *Heat kernels and Dirac operators*. Springer-Verlag, 1991.
- M.M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1704–1711, 2010.
- 7. P. Bubenik and P.T. Kim. A statistical approach to persistent homology. Homology Homotopy and Applications, 9:337–362, 2007.
- 8. A. Bueno-Orovio. Fourier embedded domain methods: periodic and  $c^{\infty}$  extension of a function defined on an irregular region to a rectangle via convolution with Gaussian kernels. *Applied Mathematics and Computation*, 183:813–818, 2006.
- G. Carlsson and F. Memoli. Persistent clustering and a theorem of J. Kleinberg. arXiv preprint arXiv:0808.2241, 2008.

- F.R.K. Chung and S.T. Yau. Eigenvalue inequalities for graphs and convex subgraphs. Communications in Analysis and Geometry, 5:575–624, 1997.
- M.K. Chung. Statistical Morphometry in Neuroanatomy. Ph.D. Thesis, McGill University, 2001. http://www.stat.wisc.edu/~mchung/papers/ thesis.pdf.
- M.K. Chung, K.M. Dalton, and R.J. Davidson. Tensor-based cortical surface morphometry via weighted spherical harmonic representation. *IEEE Transactions on Medical Imaging*, 27:1143–1151, 2008.
- 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.
- 14. M.K. Chung, R. Hartley, K.M. Dalton, and R.J. Davidson. Encoding cortical surface by spherical harmonics. *Statistica Sinica*, 18:1269–1291, 2008.
- M.K. Chung, A. Qiu, S. Seo, and H.K. Vorperian. Unified heat kernel regression for diffusion, kernel smoothing and wavelets on manifolds and its application to mandible growth modeling in CT images. *Medical Image Analysis*, 22:63–76, 2015.
- M.K. Chung, S. Robbins, K.M. Dalton, R.J. Davidson, A.L. Alexander, and A.C. Evans. Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, 25:1256–1265, 2005.
- M.K. Chung, S. Robbins, and A.C. Evans. Unified statistical approach to cortical thickness analysis. *Information Processing in Medical Imaging (IPMI)*, *Lecture Notes in Computer Science*, 3565:627–638, 2005.
- M.K. Chung and J. Taylor. Diffusion smoothing on brain surface via finite element method. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, volume 1, pages 432–435, 2004.
- M.K. Chung, Y. Wang, and G. 2018 Wu. Heat kernel smoothing in irregular image domains. *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018.
- M. Desbrun, E. Kanso, and Y. Tong. Discrete differential forms for computational modeling. In *Discrete differential geometry*, pages 287–324. Springer, 2008.
- G. Domokos. Four-dimensional symmetry. *Physical Review*, 159:1387–1403, 1967.
- 22. A. Gelb. The resolution of the Gibbs phenomenon for spherical harmonics. Mathematics of Computation, 66:699–717, 1997.
- G.M.L. Gladwell and H. Zhu. Courant's nodal line theorem and its discrete counterparts. The Quarterly Journal of Mechanics and Applied Mathematics, 55:1–15, 2002.
- L.R. Haff, P.T. Kim, J.-Y. Koo, and D.S.P. Richards. Minimax estimation for mixtures of wishart distributions. *The Annals of Statistics*, 39:3417–3440, 2011.
- A.P. Hosseinbor, M.K. Chung, C.G. Koay, S.M. Schaefer, C.M. Van Reekum, L.P. Schmitz, M. Sutterer, A.L. Alexander, and R.J. Davidson. 4D hyperspherical harmonic (HyperSPHARM) representation of surface anatomy: A

### M. K. Chung & Y. Wang

holistic treatment of multiple disconnected anatomical structures. *Medical Image Analysis*, 22:89–101, 2015.

- A.P. Hosseinbor, M.K. Chung, Y.-C. Wu, B.B. Bendlin, and A.L. Alexander. A 4D hyperspherical interpretation of q-space. *Medical Image Analysis*, 21:15–28, 2015.
- A.P. Hosseinbor, W.H. Kim, N. Adluru, A. Acharya, H.K. Vorperian, and M.K. Chung. The 4D hyperspherical diffusion wavelet: A new method for the detection of localized anatomical variation. In *International Conference* on *Medical Image Computing and Computer-Assisted Intervention*, volume 8675, pages 65–72, 2014.
- A.T. James. Calculation of zonal polynomial coefficients by use of the Laplace-Beltrami operator. *The Annals of Mathematical Statistics*, 39:1711– 1718, 1968.
- 29. E. Kreyszig. Differential Geometry. University of Toronto Press, 1959.
- B. Lévy. Laplace-Beltrami eigenfunctions towards an algorithm that "understands" geometry. In *IEEE International Conference on Shape Modeling and Applications*, page 13, 2006.
- C. Li and V. Alexiades. Time stepping for the cable equation, Part 1: Serial performance. Proceedings of Neural, Parallel & Scientific Computations, 4:241-246, 2010.
- H. Maass. Die bestimmung der dirichletreihen mit grössencharakteren zu den modulformen n-ten grades. Journal of Indian Mathematical Society, 19:1–23, 1955.
- J.E. Marsden and T.J.R. Hughes. Mathematical Foundations of Elasticity. Dover Publications, Inc., 1983.
- J. Nilsson, F. Sha, and M.I. Jordan. Regression on manifolds using kernel dimension reduction. In *Proceedings of the 24th International Conference on Machine Learning*, pages 697–704. ACM, 2007.
- 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.
- D.S.P. Richards. Applications of invariant differential operators to multivariate distribution theory. SIAM Journal on Applied Mathematics, 45:280–288, 1985.
- D.S.P. Richards. High-dimensional random matrices from the classical matrix groups, and generalized hypergeometric functions of matrix argument. Symmetry, 3:600–610, 2011.
- S. Rosenberg. The Laplacian on a Riemannian Manifold. Cambridge University Press, 1997.
- 39. B. Schölkopf and A.J. Smola. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, 2002.
- S. Seo, M.K. Chung, and H.K. Vorperian. Heat kernel smoothing using Laplace-Beltrami eigenfunctions. In Medical Image Computing and Computer-Assisted Intervention — MICCAI 2010, volume 6363 of Lecture Notes in Computer Science, pages 505–512, 2010.

- 41. J. Shawe-Taylor and N. Cristianini. *Kernel methods for pattern analysis*. Cambridge University Press, 2004.
- 42. D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending highdimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30:83–98, 2013.
- O. Stein, E. Grinspun, M. Wardetzky, and A. Jacobson. Natural boundary conditions for smoothing in geometry processing. arXiv preprint arXiv:1707.04348, 2017.
- 44. Florian Steinke and Matthias Hein. Non-parametric regression between manifolds. Advances in Neural Information Processing Systems, 2008.
- J. Sun, M. Ovsjanikov, and L. J. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Comput. Graph. Forum*, 28:1383–1392, 2009.
- 46. T. Tlusty. A relation between the multiplicity of the second eigenvalue of a graph laplacian, courants nodal line theorem and the substantial dimension of tight polyhedral surfaces. *Electrnoic Journal of Linear Algebra*, 16:315–24, 2007.
- 47. G. Wang, X. Zhang, Q. Su, J. C., J. Chen, L. Wang, Y. Ma, Q. Liu, L. Xu, J. Shi, and Y. Wang. A heat kernel based cortical thickness estimation algorithm. In *International Workshop on Multimodal Brain Image Analysis*, pages 233–245, 2013.
- G. Wang, X. Zhang, Q. Su, J. Shi, R.J. Caselli, Y. Wang, and Alzheimer's Disease NeuroImaging Initiative. A novel cortical thickness estimation method based on volumetric Laplace-Beltrami operator and heat kernel. *Medical Image Analysis*, 22:1–20, 2015.
- E. Wolfson and E.L. Schwartz. Computing minimal distances on polyhedral surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:1001–1005, 1989.
- G. Wu, Q. Wang, J. Lian, and D. Shen. Estimating the 4d respiratory lung motion by spatiotemporal registration and super-resolution image reconstruction. *Medical Physics*, 40(3):031710, 2013.
- F. Yger and A. Rakotomamonjy. Wavelet kernel learning. *Pattern Recogni* tion, 44(10-11):2614–2629, 2011.

August 29, 2019 BC: 11272 - Lecture Note Series, IMS, NUS — Review Vol. 9in x 6in 08 page 210