Discrete Heat Kernel Smoothing in Irregular Image Domains

Moo K. Chung, Yanli Wang, Guorong Wu

***Abstract*— We present the discrete version of heat kernel smoothing on graph data structure. The method is used to smooth data in an irregularly shaped domains in 3D images. New statistical properties of heat kernel smoothing are derived. As an application, we show how to filter out noisy data in the lung blood vessel trees obtained from computed tomography. The method can be further used in representing the complex vessel trees parametrically as a linear combination of basis functions and extracting the skeleton representation of the trees.**

I. Introduction

Heat kernel smoothing was originally introduced in the context of filtering out surface data defined on mesh vertices obtained from 3D medical images [6], [5]. The formulation uses the tangent space projection in approximating the heat kernel by iteratively applying Gaussian kernel with smaller bandwidth. Recently proposed spectral formulation to heat kernel smoothing [4] constructs the heat kernel analytically using the eigenfunctions of the Laplace-Beltrami (LB) operator, avoiding the need for the linear approximation used in [6], [8].

In this paper, we present the discrete version of heat kernel smoothing on graphs. Instead of Laplace-Beltrami operator, graph Laplacian is used to construct the discrete version of heat kernel smoothing. New statistical properties are derived for heat kernel smoothing that utilizes the fact heat kernel is a probability distribution. Heat kernel smoothing is used to filter out data defined on irregularly shaped domains in 3D images.

As a demonstration, the method is applied in irregularly shaped lung blood vessel trees obtained from computed tomography (CT). 3D image volumes are represented as a large 3D graph by connecting neighboring voxels in the vessel trees. Since heat kernel smoothing is analytically represented as a linear combination of Laplace eigenfunctions, it is possible to resample lung vessel trees in a higher resolution and obtain the skeleton representation of the trees for further shape analysis [11], [2].

II. Preliminary

Let $G = \{V, E\}$ be a graph with vertex set $V = \{1, 2, \cdots, p\}$ and edge set E. 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. The graph Laplacian L can then be written as $L = D - W$, where $D = (d_{ij})$ is the diagonal matrix

*The correspondence should be sent to mkchung@wisc.edu. Moo K. Chung is with University of Wisconsin, Madison, USA, Yanli Wang is with Institute of Applied Physics and Computational Mathematics, Beijing, China, Guorong Wu is University of North Carolina, Chapel Hill, US.

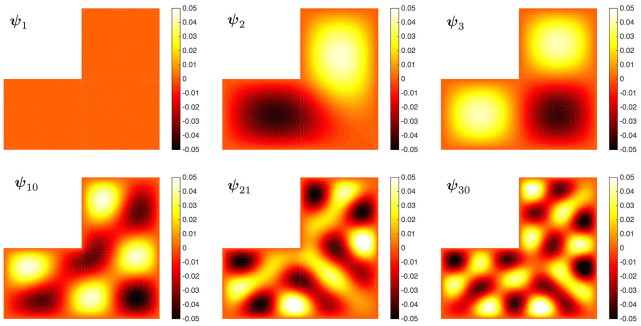

Fig. 1. First few eigenvectors of the Laplacian in a L-shaped domain.

with $d_{ii} = \sum_{j=1}^{n} w_{ij}$. For this study, we will simply use the adjacency matrix so that the edge weights w_{ij} are either 0 or 1. We have up to p number of eigenvectors $\psi_1, \psi_2, \cdots, \psi_p$ satisfying

$$L\psi_j = \lambda_j \psi_j \tag{1}$$

with $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_p$. 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 $\mathbf{1} = (1, 1, \cdots, 1)^{\mathsf{T}}$.

All other higher order eigenvalues and eigenvectors are unknown analytically and have to be computed numerically (Figure 1). Using the eigenvalues and eigenvectors, the graph Laplacian can be decomposed spectrally. From (1),

$$L\Psi = \Psi\Lambda, \tag{2}$$

where $\Psi = [\psi_1, \cdots, \psi_p]$ and Λ is the diagonal matrix with entries $\lambda_1, \cdots, \lambda_p$. Since Ψ 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.

For measurement vector $f = (f_1, \cdots, f_p)^{\mathsf{T}}$ observed at the p nodes, the discrete Fourier series expansion is given by

$$f = \sum_{j=1}^{n} \tilde{f}_j \psi_j,$$

where $\tilde{f}_j = f^{\mathsf{T}}\psi_j = \psi_j^{\mathsf{T}} f$ are Fourier coefficients.

III. Heat kernel on graphs

The *discrete heat kernel* K_σ is a positive definite symmetric matrix of size $p \times p$ given by

$$K_\sigma = \sum_{j=1}^{p} e^{-\lambda_j \sigma} \psi_j \psi_j^{\mathsf{T}}, \tag{3}$$

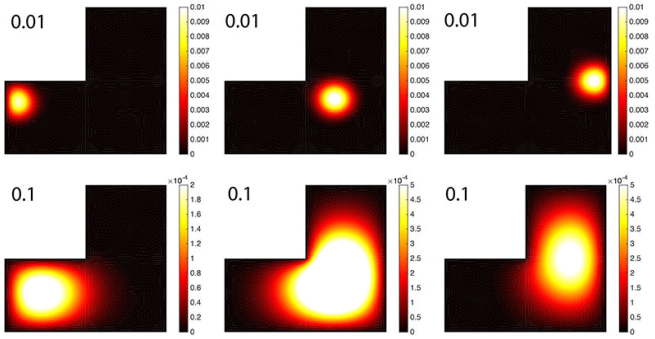

Fig. 2. Heat kernel with bandwidths $\sigma = 0.01, 0.1$. We have used degree 70 expansions but the shape is almost identical if we use higher degree expansions. The heat kernel is a probability distribution that follows the shape of the L-shaped domain.

where σ is called the bandwidth of the kernel. Figure 2 shows heat kernel with different bandwidths at a L-shaped domain. Alternately, we can write (3) as

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

where $e^{-\sigma\Lambda}$ is the matrix logarithm of Λ. When $\sigma = 0$, $K_0 = I_p$, identity matrix. When $\sigma = \infty$, by interchanging the sum and the limit, we obtain

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

K_∞ 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 [3] so that

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

Thus, K_σ 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

$$\begin{aligned} K_\sigma^2 &= \sum_{i,j=1}^{p} e^{-(\lambda_i+\lambda_j)\sigma}\psi_i\psi_i^{\mathsf{T}}\psi_j\psi_j^{\mathsf{T}} \\ &= \sum_{j=1}^{p} e^{-2\lambda_j\sigma}\psi_j\psi_j^{\mathsf{T}} = K_{2\sigma}. \end{aligned}$$

We used the orthonormality of eigenvectors. Subsequently, we have

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

IV. Heat kernel smoothing on graphs

Discrete heat kernel smoothing of measurement vector f is then defined as convolution

$$K_\sigma * f = K_\sigma f = \sum_{j=0}^{p} e^{-\lambda_j\sigma}\tilde{f}_j\psi_j, \tag{4}$$

This is the discrete analogue of heat kernel smoothing first defined in [6]. In discrete setting, the convolution $*$ is simply a matrix multiplication. Thus,

$$K_0 * f = f$$

and

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

where $\bar{f} = \sum_{j=1}^{p} f_j/p$ is the mean of signal f over every nodes. When the bandwidth is zero, we are not smoothing data. As the bandwidth increases, the smoothed signal converges to the sample mean over all nodes.

Define the l-norm of a vector $f = (f_1, \cdots, f_p)^{\mathsf{T}}$ as

$$\| f \|_l = \Big(\sum_{j=1}^{p} |f_j|^l\Big)^{1/l}.$$

The matrix ∞-norm is defined as

$$\| f \|_\infty = \max_{1\le j\le p} |f_j|.$$

Theorem 1. *Heat kernel smoothing is a contraction mapping with respect to the l-th norm, i.e.,*

$$\|K_\sigma * f\|_l^l \le \|f\|_l^l.$$

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

$$\|K_\sigma * f\|_l^l = \sum_{i=1}^{p}\sum_{j=1}^{p} |k_{ij}f_j|^l \le \sum_{j=1}^{p} |f_j|^l.$$

We used Jensen's inequality and doubly-stochastic property of heat kernel. Similarly, we can show that heat kernel smoothing is a contraction mapping with respect to the ∞-norm as well.

Theorem 1 shows that heat kernel smoothing contracts the overall size of data. This fact can be used to skeltonize the blood vessel trees.

V. Statistical properties

Often observed noisy data f on graphs is smoothed with heat kernel K_σ to increase the signal-to-noise ratio (SNR) and increases the statistical sensitivity [4]. We are interested in knowing how heat kernel smoothing will affect on the statistical properties of smoothed data.

Consider the following addictive noise model:

$$f = \mu + e, \tag{5}$$

where μ is unknown signal and ϵ is zero mean noise. Let $e = (e_1, \cdots, e_p)^{\mathsf{T}}$. Denote $\mathbb{E}$ as expectation and $\mathbb{V}$ as covariance. It is natural to assume that the noise variabilities at different nodes are identical, i.e.,

$$\mathbb{E}e_1^2 = \mathbb{E}e_2^2 = \cdots = \mathbb{E}e_p^2. \tag{6}$$

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

$$r_{ij} = \rho(d_{ij}) \tag{7}$$

for some decreasing function ρ 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 diagonals.

Noise e can be further modeled as Gaussian white noise, i.e., Brownian motion or the generalized derivatives of Wiener process, whose covariance matrix elements are Dirac-delta. For the discrete counterpart, $r_{ij} = \delta_{ij}$, where δ_{ij} is Kroneker-delta 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 $\delta_{jj} \geq \delta_{ij}$, Gaussian white noise is a special case of (7).

Once heat kernel smoothing is applied to (5), we have

$$K_\sigma * f = K_\sigma * \mu + K_\sigma * e. \tag{8}$$

We are interested in knowing how the statistical properties of model change from (5) to (8). 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

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

The variance of data will be often reduced after heat kernel smoothing in the following sense [6], [5]:

Theorem 2. *Heat kernel smoothing reduces variability, i.e.,*

$$\mathbb{V}(K_\sigma * f)_j \leq \mathbb{V} f_j$$

for all j. The subscript $_j$ indicates the j-th element of the vector.

Proof. Note

$$\mathbb{V}(K_\sigma * f)_j = \mathbb{V}(K_\sigma * e)_j = \mathbb{E}\Big(\sum_{i=1}^{p} k_{ij} e_i\Big)^2.$$

Since (k_{ij}) is doubly-stochastic, after applying Jensen's inequality, we obtain

$$\mathbb{E}\Big(\sum_{i=1}^{p} k_{ij} e_i\Big)^2 \leq \mathbb{E}\Big(\sum_{i=1}^{p} k_{ij} e_i^2\Big) = \mathbb{E} e_i^2.$$

For the last equality, we used the equality of noise variability (6). Since $\mathbb{E} f_j = \mathbb{E} e_i^2$, we proved the statement.

Theorem 2 shows that the variability of data decreases after heat kernel smoothing.

VI. Application

We applied heat kernel smoothing to the computed tomography (CT) of human lung obtained from DIR-lab dataset (`https://www.dir-lab.com`) with $1 \times 1 \times 1$ mm^3 resolution [1], [13]. The small part ($50 \times 50 \times 40$ voxels) of the CT image was used to illustrate the method and display the results better. The binary vessel segmentation was done using the multiscale Hessian filters at each voxel, as shown in Figure 3 [7], [9], [12].

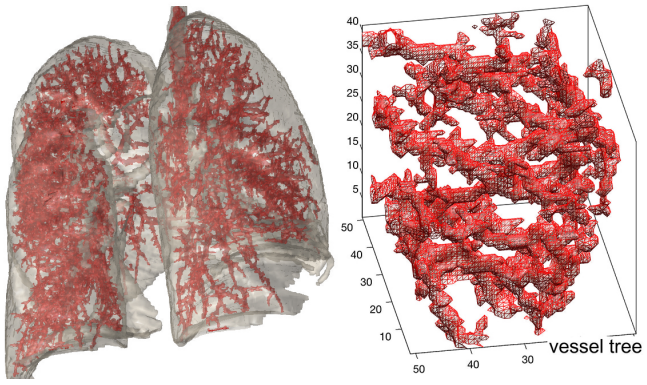

Fig. 3. The 3D reconstruction of lung vessel tree. The small part ($50 \times 50 \times 40$ voxeles) of the image is used for illustrating heat kernel smoothing.

The binary segmentation was converted into a 3D graph by taking each voxel as a node and connecting neighboring voxels. Using the 18-connected neighbor scheme, we connect two voxels only if they touch each other on their faces or edges. If voxels are only touching at their corner vertices, they are not considered as connected. Although we used the 18-connected neighbor scheme in this study, for visualization purpose only, Figure 4 displays the 6-connected neighbor scheme. This results in an adjacency matrix and the 3D graph Laplacian. Figure 4 example results in a 3D graph with $p = 6615$ nodes. The center of voxel is taken as the node coordinates. The large-scale eigenvector problem was subsequently solved using an Implicitly Restarted Arnoldi Iteration method [10]. We used 6000 eigenvectors. Note we cannot have more eigenvectors than the number of nodes.

As an illustration, we performed heat kernel smoothing on a simulated data. Gaussian noise is added to one of the coordinates. Heat kernel smoothing is performed on the noise added coordinate. Numbers in Figure 4 are kernel bandwidths. At $\sigma = 0$, heat kernel smoothing is equivalent to Fourier series expansion. Thus, we get the almost identical result. As the bandwidth increases, smoothing converges to the mean value. Each disconnected regions should converge to their own different mean values. Thus, when $\sigma = 10000$, the regions that are different colors are regions that are disconnected.

The proposed technique can be used to extract the skeleton representation of vessel trees. Here we show the proof of concept. We perform heat kernel smoothing on node coordinates with $\sigma = 1$. Then rounded off the smoothed coordinates to the nearest integers. The rounded off coordinates were used to reconstruct the binary segmentation. This gives the thick trees in Figure 5 (top left). To obtain thinner trees, the smoothed coordinates were scaled by the factor of 2, 4 and 6 times before rounding off. This had the effect of increasing the image size relative to the kernel bandwidth thus obtaining the skeleton representation of the complex blood vessel (Figure 5 clockwise from top right) [11], [2].

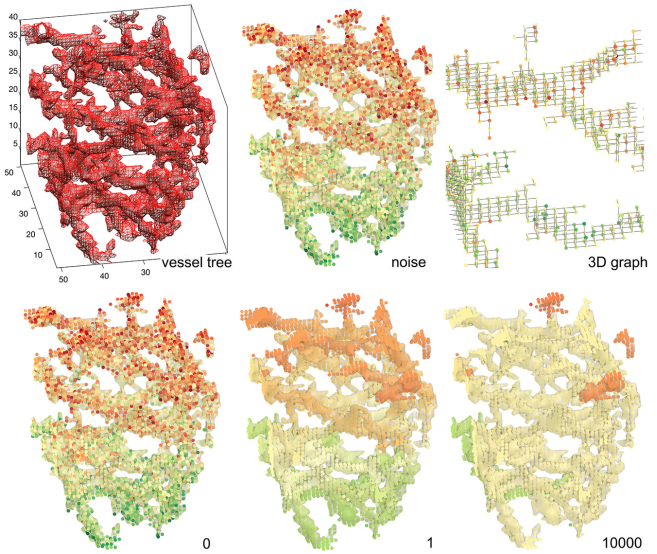

Fig. 4. From top left to right: 3D lung vessel tree. Gaussian noise is added to one of the coordinates. 3D graph constructed using 6-connected neighbors. The numbers are the kernel bandwidth σ.

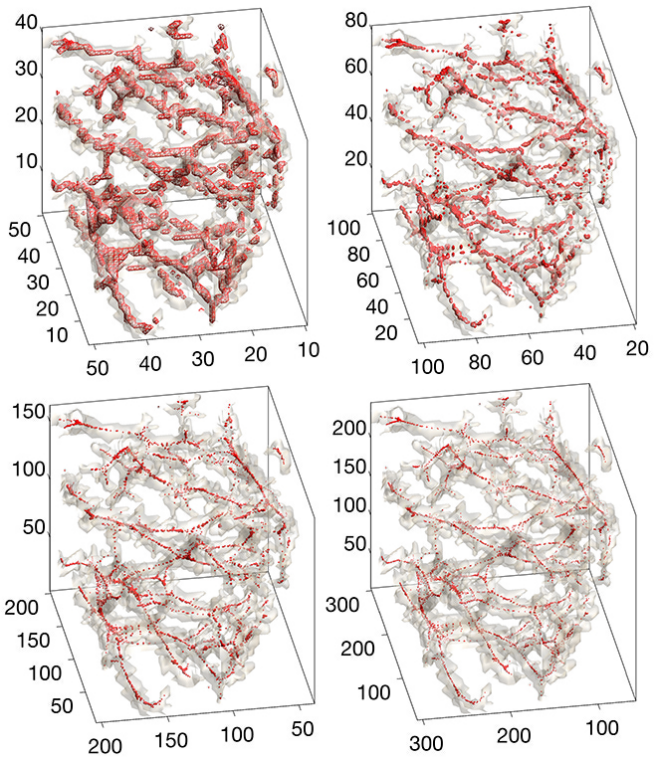

Fig. 5. The skeleton representation of vessel trees. Using the heat kernel series expansion with bandwidth $\sigma = 1$ and 6000 basis, we upsampled the binary segmentation at 2, 4, 6 times (clockwise from top right) larger than the original size (top left).

VII. Conclusion

We presented the discrete version of heat kernel smoothing and used in showing to to filter noisy data along the irregularly shaped blood vessel tree. By making the blood vessel tree in CT as a 3D graph, heat kernel smoothing was constructed as a linear combination of graph Laplacian eigenfunctions. The method is further used to effectively make the skeleton representation of the vessel trees. It may be possible to make a simpler but topologically equivalent 3D graph representation using the skeletonized blood vessel. The method can be further used in estimating the number of disconnected structures in complex data by performing heat kernel smoothing with extremely large bandwidth.

Acknowledgement

This work was supported in part by NIH grant R01 EB022856. We would like to thank Jim Ramsay of McGill University and Michelle Carey of University College Dublin for valuable discussion on solving partial differential equations on irregular domains. We would like to thank Ruth Sullivan, Michael Johnson and Michael Newton of University of Wisconsin-Madison for discussion on modeling vessel trees.

References

[1] R. Castillo, E. Castillo, R. Guerra, V.E. Johnson, T. McPhail, A.K. Garg, and T. Guerrero. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in medicine and biology*, 54:1849–1870, 2009.

[2] L. Cheng, J. De, X. Zhang, F. Lin, and H. Li. Tracing retinal blood vessels by matrix-forest theorem of directed graphs. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 626–633. Springer, 2014.

[3] F.R.K. Chung and S.T. Yau. Eigenvalue inequalities for graphs and convex subgraphs. *Communications in Analysis and Geometry*, 5:575–624, 1997.

[4] 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.

[5] 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.

[6] 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.

[7] A.F. Frangi, W.J. Niessen, K.L. Vincken, and M.A. Viergever. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 1496, pages 130–137, 1998.

[8] X. Han, J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco, M. Albert, R. Killiany, et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage*, 32:180–194, 2006.

[9] P.D. Korfiatis, C. Kalogeropoulou, A.N. Karahaliou, A.D. Kazantzi, and L.I. Costaridou. Vessel tree segmentation in presence of interstitial lung disease in MDCT. *IEEE Transactions on Information Technology in Biomedicine*, 15:214–220, 2011.

[10] R.B. Lehoucq and D.C. Sorensen. Deflation techniques for an implicitly restarted arnoldi iteration. *SIAM Journal on Matrix Analysis and Applications*, 17:789–821, 1996.

[11] L. Lindvere, R. Janik, A. Dorr, D. Chartash, B. Sahota, J.G. Sled, and B. Stefanovic. Cerebral microvascular network geometry changes in response to functional stimulation. *Neuroimage*, 71:248–259, 2013.

[12] Y. Shang, R. Deklerck, E. Nyssen, A. Markova, J. de Mey, X. Yang, and K. Sun. Vascular active contour for vessel tree segmentation. *IEEE Transactions on Biomedical Engineering*, 58:1023–1032, 2011.

[13] 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.