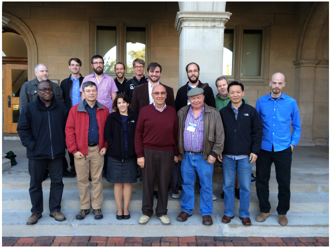
Dual Geometry of Laplacian Eigenfunctions and Graph Spatial-Spectral Analysis

Alex Cloninger

Department of Mathematics
and
Halicioğlu Data Science Institute
University of California, San Diego


Collaborators

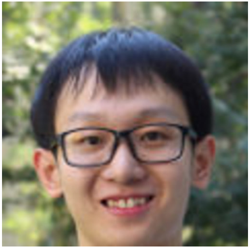
- Dual Geometry:

Stefan Steinerberger (Yale)

- Graph Wavelets:

Naoki Saito (UC Davis)

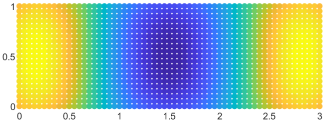
Haotian Li (UC Davis)



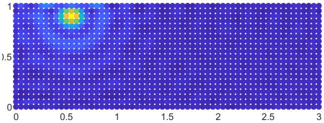
Outline

Geometric Data Representation

- In many data problems, important to create dictionaries that induce sparsity
  - Function regression / denoising
  - Combining nearby sensor time series to filter out sensor dependent information
- Consider problem of building dictionary on graph $G = (V, E, K)$
  - Similarly induced graph from point cloud and kernel similarity
- Many graph representations built in similar way to classical Fourier / wavelet literature
  - Laplacian Eigenmaps
    - Global wave-like ONB with increasing frequency
    - Belkin, Niyogi 2005
  - Spectral wavelets
    - Localized frame built from filtering LE
    - Hammond, Gribonval, Vandergheynst 2009



Eigenfunction



Spectral Wavelet

This Talk

Topic of This Talk

- “Fourier transform on graphs” story, while tempting, is more complicated than previously understood
- Relationship between eigenvectors isn’t strictly monotonic in eigenvalue

This Talk

Topic of This Talk

- “Fourier transform on graphs” story, while tempting, is more complicated than previously understood
- Relationship between eigenvectors isn’t strictly monotonic in eigenvalue

Real Topic of This Talk

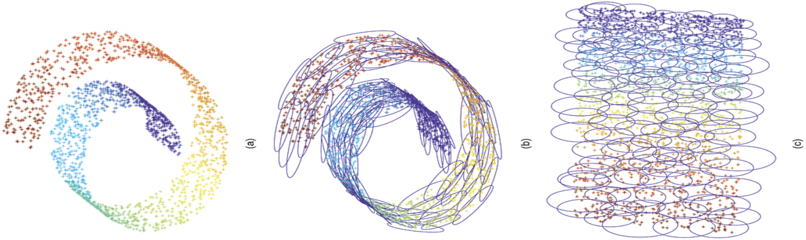
- Prove to JJB I paid attention in all the “applied harmonic analysis” classes I took here.

Kernels as Networks

- Collection of which points similar to which forms a local network graph $G = (X, E, W)$
- Graph Laplacian $\mathcal{L} := \mathcal{I} - D^{-1/2} W D^{-1/2}$, for $D_{xx} = \sum_y W_{x,y}$
- Winds up only need a few eigenfunctions to describe global characteristics

$$\mathcal{L}\phi_\ell = \lambda_\ell \phi_\ell, \qquad 0 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{N-1}$$

  - Diffusion Maps, Laplacian Eigenmaps, kPCA, Spectral Clustering
  - Filters $g(t\lambda_i)$ used to form localized wavelets



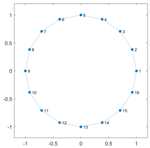
Low-dim. data — Local covering — (ϕ_1, ϕ_2) Embedding

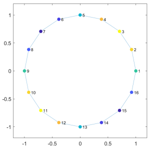
Laplacian Eigenfunctions

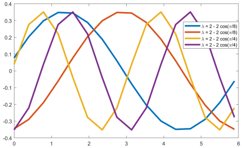
- Common to view ϕ_ℓ as Fourier basis and λ_ℓ as "frequencies" of ϕ_ℓ
  - Parallel exists for paths, cycles, bipartite graphs
  - Problematic view once move beyond simple graphs
- Fourier interpretation used to build spectral graph wavelets

$$\psi_{m,t}(x) = \sum_{\ell} g(t\lambda_\ell)\phi_\ell(x_m)\phi_\ell(x)$$

  - Filter smooth in λ_ℓ implies $\psi_{m,t}(x)$ decays quickly away from x
  - Choose g so $\sum_{t \in T} g(t\lambda) \approx 1$







Why Parallel Exists and Why Breaks Down

Connection:

- Idea exists because $\mathcal{L} \to -\Delta$, Laplacian on manifolds
  - $\Delta\ e^{-\mathrm{i}kx} = k^2 \cdot e^{-\mathrm{i}kx}$
- Parallel is convenient because easy to define low-pass filters and wavelets in Fourier space

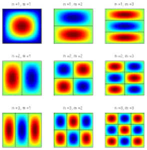
Why Parallel Exists and Why Breaks Down

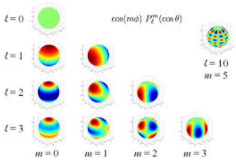
Connection:

- Idea exists because $\mathcal{L} \to -\Delta$, Laplacian on manifolds
  - $\Delta\, e^{-\mathrm{i}kx} = k^2 \cdot e^{-\mathrm{i}kx}$
- Parallel is convenient because easy to define low-pass filters and wavelets in Fourier space

However:

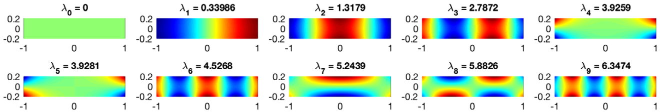
- In multiple dimensions eigenfunctions are multi-indexed according to oscillating direction (i.e. separable)
  - $F(u,v) = \int\int f(x,y) e^{-\mathrm{i}(xu+yv)} dxdy = \int\int f(x,y)\phi_{u,v}(x,y) dxdy$
- Exists entire dual geometry
  - Level-sets of equal frequency, eigenfunctions invariant in certain directions, deals with differing scales, etc.





Indexing Empirical Eigenvectors

- Graph/empirical Laplacian eigenvectors have single index λ_i regardless of dimension/structure



- Reinterpretation of multi-index is defining metric

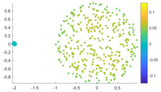
$$\rho(\phi_{u,v}, \phi_{u',v'}) = |u - u'| + |v - v'|$$

- Naive metrics on empirical eigenvectors insufficient

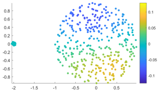
$$\|\phi_i - \phi_j\|_2 = \sqrt{2} \cdot \delta_{i,j}$$
$$\rho(\phi_i, \phi_j) = |i - j|$$

Effect of Local Scale and Number of Points

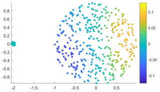
- Few points in cluster leads to most eigenfunctions concentrating in large cluster
- Geometric small cluster leads to large eigenvalue before any concentration
- If few edges connecting clusters, even fewer eigenfunctions concentrate in small cluster
  - Cloninger, Czaja 2015
- Means low-freq eigenfunctions will give rich information about large cluster only



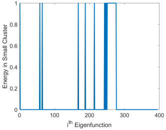
ϕ_2



ϕ_3



ϕ_4



Energy in small cluster

Larger Questions Beyond Separability

- Dual structure only readily known for small number of domains

Transform	Original domain	Transform domain
Fourier transform	$\mathbb{R}$	$\mathbb{R}$
Fourier series	$\mathbb{T}$	$\mathbb{Z}$
Discrete-time Fourier transform (DTFT)	$\mathbb{Z}$	$\mathbb{T}$
Discrete Fourier transform (DFT)	$\mathbb{Z}/(n)$	$\mathbb{Z}/(n)$

- Does there exist structure on general graph domains?
  - How do eigenfunctions on manifold organize?
  - What is dual geometry on social network?
- How do we apply this indexing?
  - Filtering
  - Wavelets / filter banks
  - Graph cuts

Outline

Local Vs. Global Correlation

Ideal model:

1. Define some non-trivial notion of distance/affinity $\alpha(\phi_i, \phi_j)$
   - Will be using pointwise products
2. Use subsequent embedding of affinity to define dual geometry on eigenvectors
   - MDS / KPCA
3. Apply clustering of some form to define indexing
   - k-means, greedy clustering, open to more ideas here

Local Vs. Global Correlation

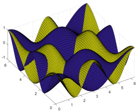
Affinity:

- Due to orthogonality, can't look at global correlation of eigenvectors
- Instead interested in notions of local similarity/correlation

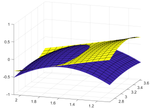
$$LC_{ij}(y) = \int M(x,y)\,(\phi_i(x) - \phi_i(y))\,\big(\phi_j(x) - \phi_j(y)\big)\,dx$$

for some local mask $M(x,y)$

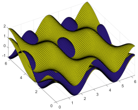
- Notion of affinity $\alpha(\phi_i, \phi_j) = \|LC_{ij}\|$
  - Characterize if ϕ_i and ϕ_j vary in same direction "most of the time"



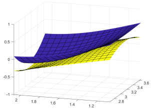
$\phi_{4,2}$ & $\phi_{2,4}$



Mean cent. $(\pi/2, \pi)$



$\phi_{4,2}$ & $\phi_{4,3}$



Mean cent. $(\pi/2, \pi)$

Intuition Behind Local Correlation

- Consider $\cos(x)$ compared to $\cos(2x)$ and $\cos(10x)$
  - Exists wavelength $\approx \pi/2$ for which most $LC_{12}(y) \neq 0$
  - Even at small bandwidth $L_{1,10}(y) \approx 0$ for large number of y
- Similarly $\cos(x_1)$ and $\cos(x_2)$ on unit square
  - $LC \approx 0$ at most (x_1, x_2)
- Questions:
  - How to define mask/bandwidth
  - How to compute efficiently
  - Proper normalization

Formalizing Relationship

- Oberved by Steinerberger in 2017 that low-energy in $\phi_\lambda \phi_\mu(x_0)$ is related to angle between at x_0 and local correlation
- In particular, making mask the heat operator yields notion of scale

Pointwise Product of Eigenfunctions

At t such that $e^{-t\lambda} + e^{-t\mu} = 1$, for heat kernel $p_t(x, y)$,

$$\left[e^{t\Delta}(\phi_\lambda \phi_\mu)\right](y) = \int p_t(x, y)\left(\phi_\lambda(x) - \phi_\lambda(y)\right)\left(\phi_\mu(x) - \phi_\mu(y)\right) dx$$

- Main relationship comes from Feynman-Kac formula
- Was considered as question about characterizing behavior of triple product $\langle \phi_i, \phi_j \phi_k \rangle$

Efficient Notion of Affinity

- Pointwise product yields much easier computation that's equivalent at diffusion time t
- Also gives notion of scale for masking function that changes with frequency
  - If mask size didn't scale, all high freq eigenvectors would cancel itself out (a la Riemann-Lebesgue lemma)
- Also want to put on the same scale to measure constructive/destructive interference
  - Can normalize by raw pointwise product
- Want geometry on data space to define geometry on the dual space through heat kernel

Eigenvector Affinity (C., Steinerberger, 2018)

We define the non-trivial eigenvector affinity for $-\Delta = \Phi\Lambda\Phi^*$ to be

$$\alpha(\phi_i, \phi_j) = \frac{\|e^{t\Delta}\phi_i\phi_j\|_2}{\|\phi_i\phi_j\|_2 + \epsilon} \text{ for } e^{-t\Lambda_i} + e^{-t\Lambda_j} = 1.$$

Landscape of Eigenfunctions

Embedding:

- Given $\alpha : \Phi \times \Phi \to [0, 1]$, need low-dim embedding
- Use simple KPCA of α

$$\alpha = V \Sigma V^*, \qquad V = \begin{bmatrix} v_1, & v_2, & \dots & v_k \end{bmatrix}$$

- Embedding $\begin{bmatrix} v_1, & v_2, & v_3 \end{bmatrix}$ captures relative relationships

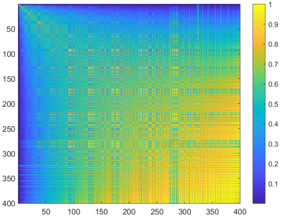
Parallel Work:

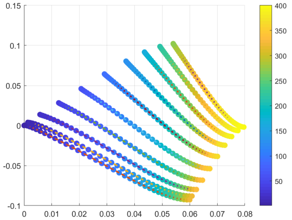
- Saito (2018) considers similar question of eig organization using ramified optimal transport on graph
  - Only defines $d(|\phi_i|, |\phi_j|)$ and slower to compute
  - Natural when eigenvectors are highly localized/disjoint

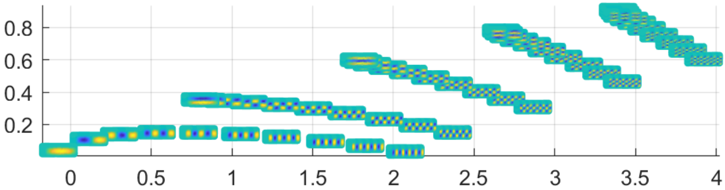
Recovery of Separable Eigenfunction Indexing

Rectangular region $[0,4] \times [0,1]$

- Eigenvectors $\sin(m\pi x)\sin(n\pi y)$ and eigenvalues $\frac{m^2}{16} + \frac{n^2}{1}$



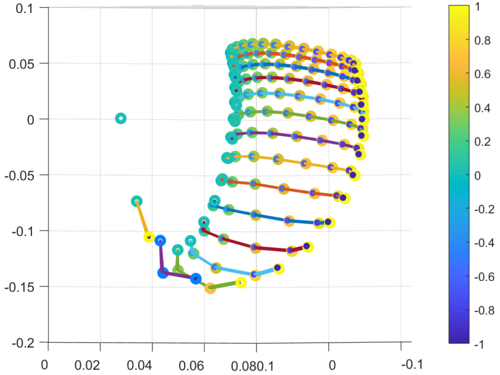




Spherical Harmonics

$$Y_\ell^m(\theta,\phi) \text{ such that } \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} Y_\ell^m Y_{\ell'}^{m'} \sin(\theta) d\phi d\theta = \delta_{m.m'}\delta_{\ell,\ell'}, \quad -m \le \ell \le m$$

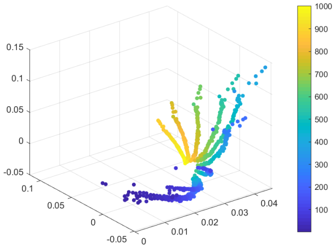
Harmonics are oriented according to (θ, ϕ), so no issue of rotational invariance



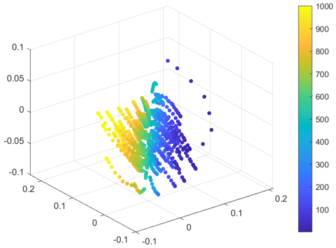
General Cartesian Product Duals

Empirical eigenvectors of graph Laplacian on Cartesian product domains for:

- $X \sim \mathcal{N}(0, \sigma^2 I_d)$ for $\sigma = 0.1$ and 100 points
- $Y \subset [0, 1]$ for 10 equi-spaced grid points
- A being adjacency matrix of an Erdos-Reyni graph



Eigs of L on $X \times Y$

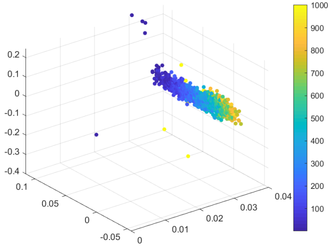


Eigs of $I - \left(A .^{*} e^{-\|x_i - x_j\|^2/\epsilon}\right)$

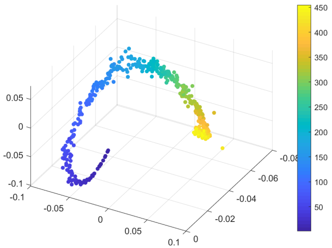
Chaotic Domains and Random Networks

Lack of structure is also captured

- Erdos-Reyni graph won't have expected structure because node neighborhood has exponential growth
- Semicircle capped rectangle (billiards domain) lacks eigenvector structure by ergodic theory (quantum chaos)



Unnormalized Erdos-Reyni Graph $p = 0.2$



Billiards domain

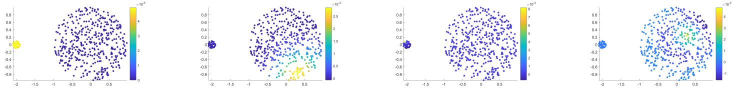
Outline

Utilizing Eigenvector Dual Geometry

- Recent work on eigenvector dual applications with Saito and Li
- Applications in spectral graph wavelet literature (Vanderghyst, et al)
  - Ideas inform modern graph CNN algorithms as well
  - Revolve around Fourier/Laplacian parallel

$$\psi_{m,t}(x) = \sum_{\ell} g(t\lambda_\ell)\phi_\ell(x_m)\phi_\ell(x)$$

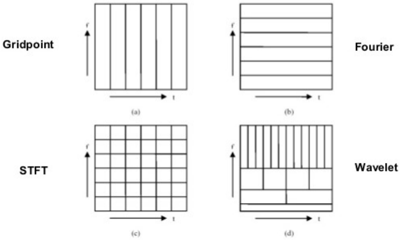
  - Problem is wavelets are inherently isotropic and use same filters $\forall x_m$



$t = 1$ $t = 1$ $t = 5$ $t = 5$

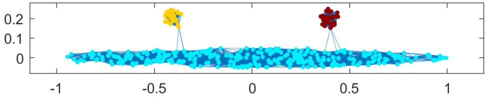
Graph Spatial-Spectral Analysis

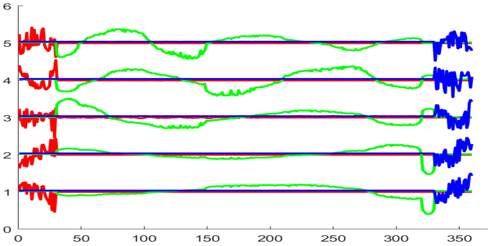
- Motivates need to construct a time-frequency tiling for nodes on graphs and their dual space
  - Relationship between nodes is more complex than path graph on time
  - Relationship between eigenfunctions is more complex than path graph on frequency
- Main problems
  - Each domain is multidimensional
  - Eigenfunction localization
  - Local correlations behave differently in different regions of network
- Basic version using Fiedler vector and eigenvalue for visualization (Ortega, et al, 2019)



Localized Eigenfunction Organization

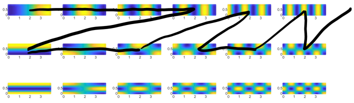
- Can split nodes via spectral clustering into K clusters $\{W_k\}_{k=1}^K$
  - Can also build hierarchical tree from iterative k-means
- Partial node affinity $\alpha(\phi_i, \phi_j; W_k)$ on each cluster
  - Non-normalized local correlation affinity using heat kernel and eigenfunctions restricted to $W_k \subset V$
- Allows for natural organization on each region separately



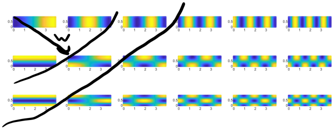


Dual Space Filters

- Graph eigenvectors give (local) similarity $\alpha^{(k)} \in \mathbb{R}^{N\times N}$ on W_k
  - Each row $\alpha^{(k)}_{i,\cdot}$ yields potential filter
  - On path graph, reduces to function of eigenvalues (indices) only
- Filter $F^{(t)}_{i,k}[j] = \dfrac{\left(\alpha^{(k)}_{i,j}\right)^{1/t}}{\sum_\ell \left(\alpha^{(k)}_{\ell,j}\right)^{1/t}}$
  - Goes to constant across spectrum as $t \to \infty$
  - Goes to indicator at $j = i$ as $t \to 0$
- $\Psi^{(t)}_{i,k} = \Phi \cdot \mathrm{diag}(F^{(t)}_{i,k}) \cdot \Phi^*$
  - Wavelet $\psi^{(t)}_{i,j,k}$ is row of $\Psi^{(t)}_{i,k}$ centered at $j \in W_k$



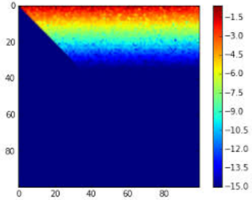
$g(t\lambda_i)$



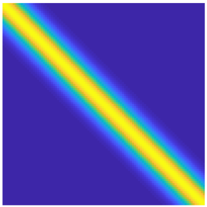
$F^{(t)}_i$

Reducing Redundancy Through QR

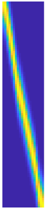
- With no reduction, there are N filters per cluster and $|W_k|$ wavelets per filter
  - Both α_k and $\Phi F_{i,k} \Phi^*$ are low rank
- Rank revealing QR with pivoting to select “prototypical points” (Chan 1990, Rokhlin 2005)
  - Low-rank, symmetric A
  - $QR = AP$ for permutation matrix P
  - Keep columns of AP such that $R_{jj} > \tau \cdot R_{11}$
  - Correspond to equivalent small set of columns E of A s.t. $\|A_{\cdot,E} A^*_{\cdot,E} - A^2\| < \tau^2$



R



α



$\alpha[:, E]$

Natural Wavelet Frame

Frame Bound (C., Li, Saito, 2019)

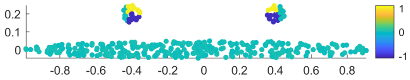
Dictionary $\{\psi_{i,j,k}^{(t)}\}_{i \in E_{\alpha_k}, j \in E_{W_k}}^{k \in \mathbb{Z}_K}$ is a frame with diagonal frame operator such that:

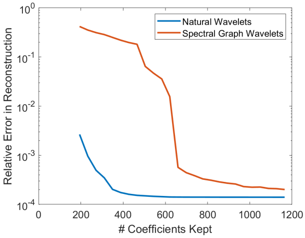
- if point sampled from smooth manifold with global eigenfunctions, $S = K \cdot I$,
- if eigenfunction localization exists, $S_{jj} = c_j$ for $j \in W_k$ where c_j depends on $\sum_i \sum_k F_{i,k}[j]$

Outline

Clustered Data

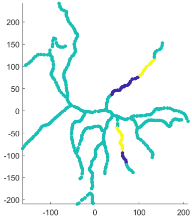
- Sparsely connected clustered graph with significantly larger/denser cluster
  - Most eigenfunctions concentrate on one cluster
  - Generic spectral wavelets don't scale for sparse representation on small clusters

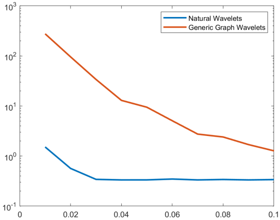




Neuronal Data

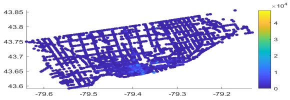
- Scan of neuron dendrite
- Eigenfunctions quickly localize on branches
  - Eigenfunctions with eigenvalue above 4 concentrate only at junctions (Saito 2011)
- Eigenfunction ordering by eigenvalue depends on length of branch



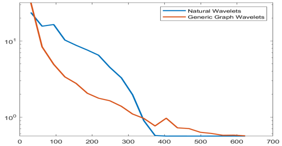


Traffic Data

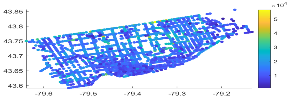
- Nodes at intersections of roads in Toronto
- No clear cluster structure, though eigenfunctions still localize
  - Low oscillations inside downtown subgraph are higher frequency than in surrounding areas
- Ordering still highly location dependent



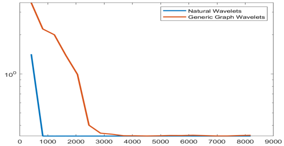
Density of People



Reconstruction MSE



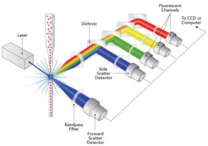
Density of Vehicles

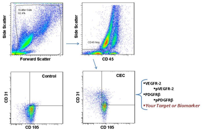


Reconstruction MSE

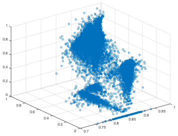
Flow Cytometry

- Flow cytometry: each patient is represented by 9D point cloud of cells

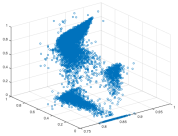




- Used to tell if people have blood disease
  - Medical test is to look at every 2D slice



Healthy

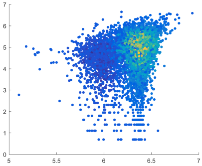


AML

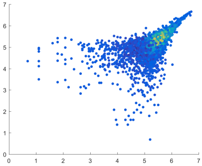
Interpretability Using Coefficients

Wavelet Application:

- Pool healthy and sick, and build network on cells
- Express cell label as function in terms of natural graph wavelets
- Examine reconstruction of largest wavelet coefficients
  - Denoise label function with low resolution wavelets that have large coefficient
- Creates function on point cloud of maximum deviation between healthy and sick cells



2D Slice Witness



2D Slice Witness

Conclusions

- Kernel/Laplacian eigenfunctions aren't like PCA
  - Don't divide into directions with independent information
  - Capable of overrepresenting certain large variance directions at expense of small scale
- Detecting relationships between eigenfunctions yields more powerful techniques while still representing geometry
- Parallel to multi-dimensional Fourier leads to new insights from harmonic analysis
- Localizing the behavior leads to appropriate scale in different places
- Using global eigenfunctions maintains smoothness across cut boundaries

Thank you!

HAPPY BIRTHDAY, JOHN!