The Mathematics and Statistics of Manifold Learning

Tyrus Berry
Dept. of Mathematical Sciences, GMU

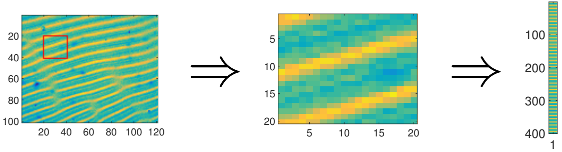
Statistics Seminar
UW-Madison
Oct. 9, 2019

Supported by NSF-DMS

ANALYSIS OF POINT CLOUDS

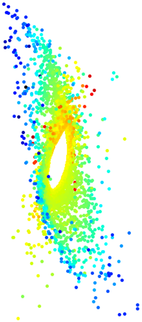
- Data lie in $\mathbb{R}^m$ for large $m \Rightarrow$ Curse-of-dimensionality
- Data may be sampled from nearly singular measures
- Geometric prior: Points lie near smooth manifold $\mathcal{M} \subset \mathbb{R}^m$
- Curse depends on the dimension $d < m$ of $\mathcal{M}$
- Goal: Learn/represent $\mathcal{M}$ with statistical error bounds

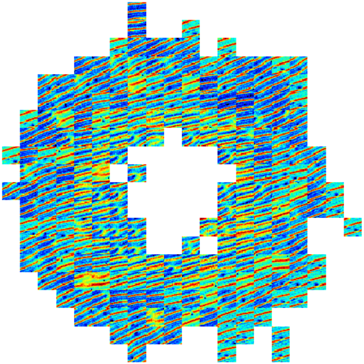
FINDING HIDDEN STRUCTURE IN DATA



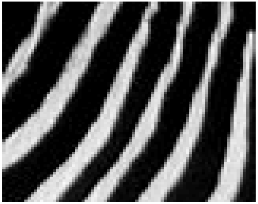
$\in \mathbb{R}^{400}$

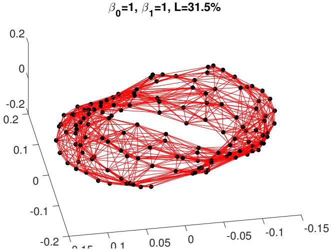
The sub-image geometry:

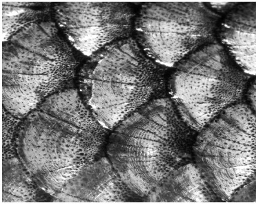


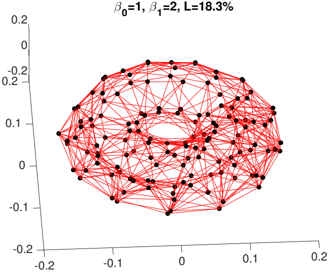


Example: Manifolds of subimages









DENSITY ESTIMATION ON $\mathbb{R}^m$

- Goal: Estimate density $p(x)$ from random variables $X_i \sim p$
- Kernel density estimation on $\mathbb{R}^m$ dates from the 1950's

$$p_{h,N}(x) \equiv \frac{1}{m_0 h^m N} \sum_{i=1}^{N} K\left(\frac{||x - X_i||}{h}\right) \qquad m_0 = \int_{\mathbb{R}^m} K(||z||)\, dz$$

- Theorem: $p_{h,N}(x)$ is a consistent estimator of $p(x)$ with
- Bias: $\mathbb{E}\left[p_{h,N}(x) - p(x)\right] = \mathcal{O}(h^2)$ and
- Variance: $\mathbb{E}\left[(p_{h,N}(x) - p(x))^2\right] = \mathcal{O}\left(\frac{h^{-m}}{N} p(x)\right)$.

Density estimation on $\mathbb{R}^m$

Finding the bias:

$$\begin{aligned}
\mathbb{E}[p_{h,N}(x)] &= \frac{1}{m_0 h^m N}\mathbb{E}\left[\sum_{i=1}^{N} K\left(\frac{||x-X_i||}{h}\right)\right] \\
&= \frac{1}{m_0 h^m}\int_{\mathbb{R}^m} K\left(\frac{||x-y||}{h}\right) p(y)\,dy \\
(\text{decay of } K) &= \frac{1}{m_0 h^m}\int_{||x-y||<h^\alpha} K\left(\frac{||x-y||}{h}\right) p(y)\,dy \\
\left(z = \frac{y-x}{h}\right) &= \frac{1}{m_0}\int_{||z||<h^{\alpha-1}} K(||z||)\, p(x+hz)\,dz \\
(\text{Taylor}) &= \frac{1}{m_0}\int_{||z||<h^{\alpha-1}} K(||z||)\left(p(x) + Dp(x)hz + \mathcal{O}(h^2||z||^2)\right) dz \\
(\text{symmetry}) &= \frac{1}{m_0}\int_{||z||<h^{\alpha-1}} K(||z||)\left(p(x) + \mathcal{O}(h^2||z||^2)\right)\ dz \\
(\alpha < 1) &= p(x)\frac{1}{m_0}\int_{\mathbb{R}^m} K(||z||)\ dz + \mathcal{O}(h^2) = p(x) + \mathcal{O}(h^2)
\end{aligned}$$

Random variables on manifolds

- A smooth embedded manifold $\mathcal{M} \subset \mathbb{R}^m$ inherits:
  - A Riemannian metric $g_x(v, w)$ (defines geometry)
  - A volume form $dV(x) = \sqrt{\det(g_x)}$
  - $\mathrm{vol}(\mathcal{M}) = \int_{x\in\mathcal{M}} 1 \, dV(x)$
- Data X_i are sampled from p supported on $\mathcal{M} \subset \mathbb{R}^m$
- Expectation integrates over a singular density:

$$\mathbb{E}_p[K(||x - X_i||)] = \int_{\mathbb{R}^m} K(||x-y||) dp(y) = \int_{y\in\mathcal{M}} K(||x-y||) q(y)\, dV(y)$$

- On manifolds we will estimate $q(x)$ where $q\, dV = dp$

Density estimation on manifolds $\mathcal{M} \subset \mathbb{R}^m$

- For data X_i sampled from $q\,dV$ on $\mathcal{M}$
- Use a kernel density estimator (KDE) of form

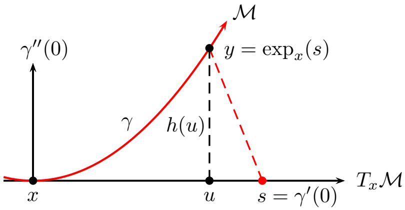
$$q_{h,N}(x) \equiv \frac{1}{N m_0 h^d} \sum_{i=1}^{N} K\left(\frac{||x - X_i||_{\mathbb{R}^n}}{h}\right), \qquad m_0 = \int_{\mathbb{R}^d} K(||z||)\, dz$$

- The expectation of the estimator is:

$$\mathbb{E}[q_{h,N}(x)] = \frac{1}{m_0 h^d N} \mathbb{E}\left[\sum_{i=1}^{N} K\left(\frac{||x - X_i||}{h}\right)\right] = \frac{1}{m_0 h^d} \int_{y \in \mathcal{M}} K\left(\frac{||x - y||}{h}\right) q(y)\, dV$$

INTEGRATING ON MANIFOLDS

The exponential map changes variables to tangent space:



- Distances preserved: $||y - x||^2 = ||s||^2 + \mathcal{O}(s_i^4)$
- Natural volume element: $dV(y) = ds$
- Taylor: $q(\exp_x(s)) = q(x) + \nabla q(x) \cdot s + \mathcal{O}(s_i^2)$

Density estimation on manifold $\mathcal{M} \subset \mathbb{R}^m$

Let $\exp_x : T_x\mathcal{M} \to U \subset \mathcal{M}$

$$\mathbb{E}[q_{h,N}(x)] = \frac{1}{m_0 h^d} \int_{||x-y||<h^\alpha} K\left(\frac{||x-y||}{h}\right) q(y)\, dV(y)$$

$$(y = \exp_x(hs)) = \frac{1}{m_0 h^d} \int_{||s||<h^\alpha} K\left(\sqrt{||s||^2 + \mathcal{O}(h^2 s_i^4)}\right) q(\exp_x(s))\, ds$$

$$(\text{Taylor}) = \frac{1}{m_0} \int_{||s||<h^{\alpha-1}} \left(K(||s||) + \mathcal{O}(h^2 s_i^4) K'(||s||)/||s||\right) \cdot \left(q(x) + \nabla q(x) hs + \mathcal{O}(h^2||s||^2)\right) ds$$

$$(\text{symmetry}) = \frac{1}{m_0} \int_{||z||<h^{\alpha-1}} K(||s||)\, q(x) + \mathcal{O}(h^2||s||^2)\, ds$$

$$(\alpha < 1) = q(x) \frac{1}{m_0} \int_{\mathbb{R}^d} K(||s||)\, ds + \mathcal{O}(h^2) = q(x) + \mathcal{O}(h^2)$$

DENSITY ESTIMATION ON MANIFOLDS $\mathcal{M} \subset \mathbb{R}^m$

Using the KDE:

$$q_{h,N}(x) \equiv \frac{1}{N m_0 h^d} \sum_{i=1}^{N} K\left(\frac{||x - X_i||_{\mathbb{R}^m}}{h}\right), \qquad m_0 = \int_{\mathbb{R}^d} K\left(||z||\right)\, dz$$

For a d-dimensional manifold without boundary:

Theorem. $q_{h,N}(x)$ is a consistent estimator of $q(x)$ with

- Bias: $\mathbb{E}\left[q_{h,N}(x) - q(x)\right] = \mathcal{O}(h^2)$
- Variance: $\mathbb{E}\left[(q_{h,N}(x) - q(x))^2\right] = \mathcal{O}\left(\frac{h^{-d}}{N} q(x)\right)$

Pelletier (2005), Hein (2006), Ozakin-Gray (2009), Kim-Park (2013)

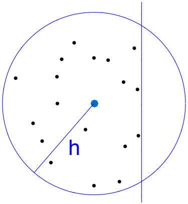
Boundary KDE

We saw that KDE extends to manifolds:

$$q_{h,N}(x) \equiv \frac{1}{N m_0 h^d} \sum_{i=1}^{N} K\left(\frac{||x - X_i||}{h}\right)$$

What about manifolds with boundary?

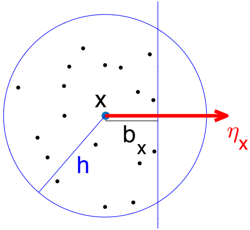
- KDE is not consistent at the boundary - normalization is wrong.
- Previous work only in $\mathbb{R}, \mathbb{R}^2$ and assumes known boundary
- Unknown boundary ???



Problem near the boundary

Suppose we can locate the boundary

and b_x is distance from x to boundary



Integration is truncated in direction of boundary, η_x

$$m_0^{\partial}(x) = \int_{\mathbb{R}^{m-1}} \int_{-\infty}^{b_x/h} K\left(\sqrt{||z_\perp||^2 + z_{||}^2}\right) \, dz_{||} dz_\perp$$

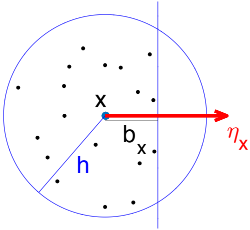
If K is Gaussian kernel then:

$$m_0^{\partial}(x) = (1 + \mathrm{erf}(b_x/h))/2$$

Locating the boundary

Boundary Direction Estimator (BDE)

$$\mu(x) \equiv \frac{1}{Nh^d} \sum_{i=1}^{N} K\left(\frac{||x - X_i||}{h}\right)\left(\frac{X_i - x}{h}\right)$$



We prove (boundary normal coordinates):

$$\mathbb{E}[\mu(x)] = -\vec{\eta}_x m_1^{\partial}(x) q(x) + \mathcal{O}(h)$$

If K is Gaussian then:

$$m_1^{\partial}(x) = \frac{\exp(-b_x^2/h^2)}{2\sqrt{\pi}}$$

Finding b_x

First, compute the standard KDE

$$\mathbb{E}[q_{h,N}(x)] = m_0^{\partial}(x)q(x) + \mathcal{O}(h)$$

Next, compute the Boundary Direction Estimator (BDE)

$$\mathbb{E}[\mu(x)] = -\vec{\eta}_x m_1^{\partial}(x)q(x) + \mathcal{O}(h)$$

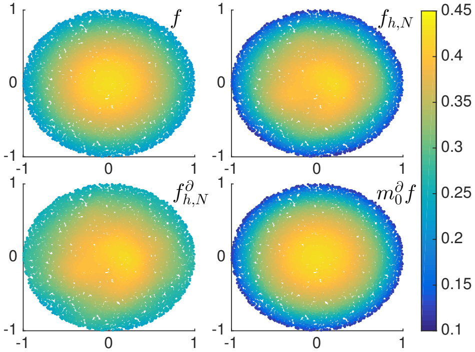
Divide to cancel the unknown $q(x)$:

$$\mathbb{E}\left[\frac{q_{h,N}(x)}{||\mu(x)||}\right] = \frac{\sqrt{\pi}(1 + \mathrm{erf}\,(b_x/h))}{\exp(-b_x{}^2/h^2)} + \mathcal{O}(h)$$

Use Newton's method to solve for b_x

Now we can compute the correct normalization!

Boundary KDE on Disc



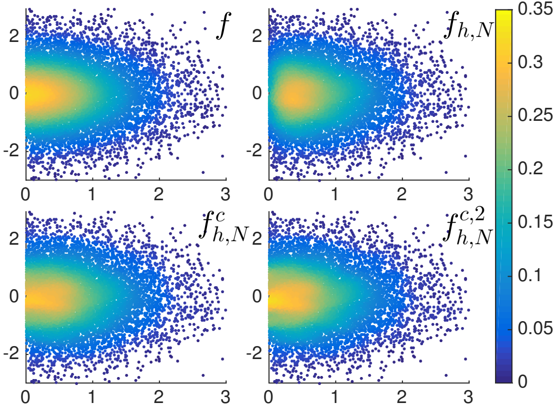
$f(x) =$ example distribution

$f_{h,N}(x) =$ standard KDE

$f^{\partial}_{h,N}(x) =$ corrected KDE using $m^{\partial}_0(x)$

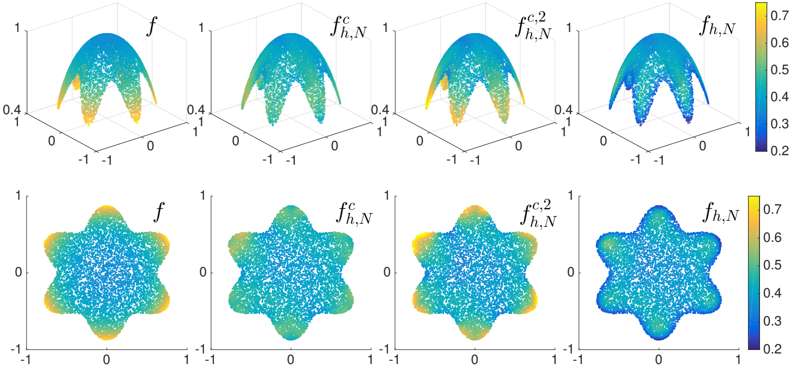
Boundary KDE on Half Gaussian

We also derive higher order estimator $f^{c,2}_{h,N}$ (extrapolation).



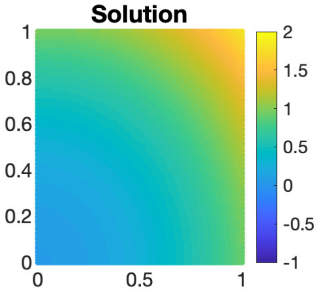
DENSITY ESTIMATION ON MANIFOLDS

Another example



BOUNDARY CONDITIONS (IN PROGRESS)

With Ryan Vaughn (graduating this Spring) and Harbir Antil



Solving $\Delta u = f$ with $\nabla u = g$ requires estimator for $\int_{\partial\mathcal{M}} g \, d_{\partial\mathcal{M}} V$

Manifold learning

- **Goal**: Represent all the information about a manifold
- Riemannian metric, g, contains all geometric information
- Laplace-Beltrami operator, Δ, is equivalent to g
- Manifold learning $\Leftrightarrow$ Estimating Laplace-Beltrami
- **Caveat:** Cannot easily answer all questions about manifold

What is Manifold Learning?

- **Manifold learning $\Leftrightarrow$ Estimating Laplace Operator**
- Euclidean space:
  - Eigenfunctions of Δ are $e^{i\vec{\omega}\cdot\vec{x}}$
  - Basis for Fourier transform
- Unit circle:
  - Eigenfunctions of Δ are $e^{ik\theta}$
  - Basis for Fourier series
- **Key Fact**: Eigenfunctions of Δ give the smoothest basis for square integrable functions on any manifold.

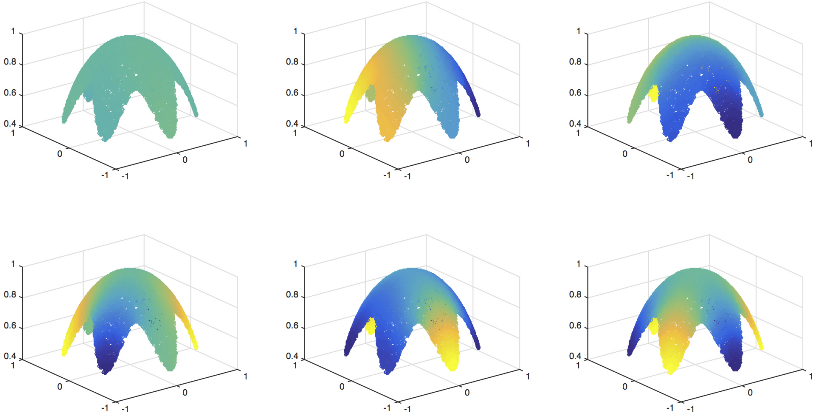
WHY THE LAPLACIAN?

- **Manifold learning $\Leftrightarrow$ Estimating Laplace-Beltrami**
- Eigenfunctions $\Delta\varphi_i = \lambda_i\varphi_i$ orthonormal basis for $L^2(\mathcal{M})$
- Smoothest functions: φ_i minimizes the functional

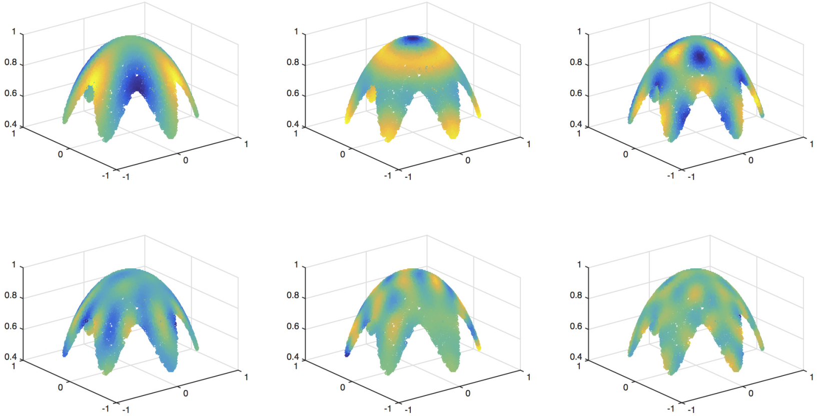
$$\lambda_i = \min_{\substack{f\perp\varphi_k \\ k=1,\ldots,i-1}} \left\{ \frac{\int_{\mathcal{M}} ||\nabla f||^2 \, dV}{\int_{\mathcal{M}} |f|^2 \, dV} \right\}$$

- Eigenfunctions of Δ are custom Fourier basis
  - Smoothest orthonormal basis for $L^2(\mathcal{M})$
  - Can be used to define wavelets
  - Define the Hilbert/Sobolev spaces on $\mathcal{M}$

Harmonic Analysis on Manifolds/Data Sets



Harmonic Analysis on Manifolds/Data Sets



Manifold learning via operator estimation

- Manifold learning $\Leftrightarrow$ Estimating Laplace-Beltrami
- To find Δ we dig deeper into the Taylor expansion:

$$f(\exp_x(s)) = f(x) + \nabla f(x) \cdot s + \frac{1}{2} s^\top H(f)(x) s$$

- Integrate over a symmetric region and odd terms drop out

$$\int_{||s||<h^{\gamma-1}} K(||s||) f(\exp_x(hs))\, ds = m_0 f(x) + \frac{h^2}{2} m_2 \Delta f(x) + \mathcal{O}(h^4)$$

Matrices as Integral Operators

- Functions are represented as vectors $\vec{f}_i = f(x_i)$
- A kernel matrix $K_{ij} = K(x_i, x_j)$ represents an operator

$$\frac{1}{N}\left(K\vec{f}\right)_i = \frac{1}{N}\sum_j K(x_i, x_j) f(x_j) \to \int_{\mathcal{M}} K(x_i, y) f(y) q(y)\, dV(y)$$

- Diagonal matrix: $D_{ii} = N^{-1}\sum_j K_{ij} = N^{-1}K\vec{1}$
- Graph Laplacian matrix: $L = \frac{1}{mh^2}\left(D^{-1}K - I\right)$
- Then $\left(L\vec{f}\right)_i = \Delta f(x_i) + \mathcal{O}(h^2)$
- This says that L is a pointwise consistent estimator of Δ

Belkin and Niyogi (2005)

DIFFUSION MAPS: ALLOWING ARBITRARY SAMPLING

- For $X_i \sim q\, dV$ on $\mathcal{M}$
- Define $K_{ij} = K\left(\frac{||x_i - x_j||}{h}\right)$ and $D_i = \sum_j K_{ij}$
- Right normalization: $\hat{K}_{ij} = K_{ij} D_j^{-1}$ and $\hat{D}_i = \sum_j \hat{K}_{ij}$
- Left normalization: $\tilde{K}_{ij} = \hat{D}_i^{-1} \hat{K}_{ij}$ and finally $L = \frac{\tilde{K} - I}{mh^2}$
- **Theorem:** L is a consistent pointwise estimator of Δ
- **Bias:** $\mathbb{E}[(L\vec{f})_i - \Delta f(x_i)] = \mathcal{O}(h^2)$
- **Variance:** $\mathbb{E}[((L\vec{f})_i - \Delta f(x_i))^2] = \mathcal{O}\left(\frac{||\nabla f(x_i)||^2 q(x_i)^{3-4d}}{N^{1/2} h^{2+d}}\right)$

Hein, Audibert, and von Luxburg (2005, 2007), Coifman and Lafon (2006)

Singer (2006), B and Harlim (2015)

HUGE PROBLEM!

- **Variance:** $\mathbb{E}[((L\vec{f})_i - \Delta f(x_i))^2] = \mathcal{O}\left(\frac{||\nabla f(x_i)||^2 q(x_i)^{3-4d}}{Nh^{2+d}}\right)$
- Exponent $3 - 4d < 0$
- If q is not bounded away from zero the variance blows up!
- **Solution:** Variable bandwidth (balloon estimator)
- For KDE variable bandwidth has little advantage:

"Nevertheless, for densities of the usual shapes, balloon estimators are mostly hot air."

-David W. Scott

- For Operator estimation variable bandwidth is important

B and Harlim (2015)

Variable Bandwidth Kernels

Repeating the Diffusion Maps construction with the kernel:

$$K_{h,\beta}(x,y) = h\left(\frac{||x-y||}{hq(x)^{\beta/2}q(y)^{\beta/2}}\right)$$

We find:

$$L_{h,\beta}f = \Delta f + \mathcal{O}\left(h^2, \frac{||\nabla f||q^{-c_2}}{\sqrt{N}h^{1+d/2}}\right)$$

- Variance determined by: $c_2 = 2d - 3/2 + d^2\beta + 3d\beta/2 - \beta$
- $c_2 < 0$ requires $\beta < \frac{-2d+3/2}{d^2+3d/2-1} \approx \frac{-2}{d}$

B and Harlim (2015)

Summary of Operator Estimation

- Kernel matrices estimate operators
- Diffusion maps normalized Laplacian $L = \frac{1}{mh^2}\left(\hat{D}^{-1}\hat{K} - I\right)$ is pointwise consistent for any sampling density
- Variance is unbounded as density decreases
- More data $\Rightarrow$ More points with low density $\Rightarrow$ Higher error!
- Variable bandwidth gives control over variance

CONTINUOUS K-NEAREST NEIGHBORS

- Let x_k denote the kth nearest neighbor of x
- Construct graph by:

CkNN: Edge between the points x, y if $\frac{||x-y||}{\sqrt{||x-x_k||\ ||y-y_k||}} < h$

- Density: $||x - x_k|| \propto q(x)^{-1/d}$ so $\beta = -1/d$
- This is a variable bandwidth kernel with $K(t) = 1_{\{t<1\}}$ so

$$K\left(\frac{||x-y||}{h\sqrt{q(x)^{-1/d}q(y)^{-1/d}}}\right) = 1_{\left\{\frac{||x-y||}{\sqrt{||x-x_k||\,||y-y_k||}}<h\right\}}$$

What about the other Riemannian structure?

- A Riemannian manifold has an exterior calculus:
  - Calculus of tensors and differential forms
  - Built entirely from the Riemannian metric $g \Leftrightarrow \Delta$
  - Formulates the generalization of the FTC (Stokes' Thm)
  - Can construct Laplacians on k-forms, Δ_k
  - Eigenforms of Δ_k are smoothest basis for k-forms
- **Question:** Given only the eigenfunctions of the Laplacian how can we construct the rest of the exterior calculus?

WHAT ABOUT THE OTHER RIEMANNIAN STRUCTURE?

- **Good News:** Laplacian $\Leftrightarrow$ Riemannian metric

$$g(\nabla f, \nabla h) = \nabla f \cdot \nabla h = \frac{1}{2}(f\Delta h + h\Delta f - \Delta(fh))$$

- Let $v, w \in T_x\mathcal{M}$, there exists $f_1, ..., f_d$ such that $\nabla f_1, ..., \nabla f_d$ span $T_x\mathcal{M}$ and

$$g(v, w) = v \cdot w = \sum_{ij} v_i w_j \nabla f_i \cdot \nabla f_j$$

- **Bad News:** There may be no $f_1, ..., f_d$ that work for all x
- Hairy Ball Thm: Every smooth vector field on S^2 must vanish: at these points the gradients do not span $T_x\mathcal{M}$.

HOW CAN WE USE THE LAPLACIAN EIGENFUNCTIONS?

- Cannot find $\nabla f_1, ..., \nabla f_d$ basis for all $T_x\mathcal{M}$
- **Whitney:** We can find $\nabla f_1, ..., \nabla f_{2d}$ span all $T_x\mathcal{M}$
- **Thm[1]:** $\exists J$ such that $\nabla\varphi_1, ..., \nabla\varphi_J$ span all $T_x\mathcal{M}$
- Representing vector fields in a frame (overcomplete set)
  - Let $v(x) \in T_x\mathcal{M}$ be a smooth vector field
  - Then $v(x) = \sum_{j=1}^{J} c_j(x)\nabla\varphi_j(x)$ where $c_j(x)$ are smooth
  - So $c_j(x) = \sum_{i=1}^{\infty} c_{ij}\varphi_i(x)$
  - Finally $v = \sum_{i,j} c_{ij}\varphi_i\nabla\varphi_j$ (not uniquely)

[1] J. Portegies, Embeddings of Riemannian Manifolds with Heat Kernels and Eigenfunctions. (2014).

HOW CAN WE USE THE LAPLACIAN EIGENFUNCTIONS?

- **Thm (Berry & Giannakis)** Let φ_i be the eigenfunctions of the Laplacian then $\{\varphi_i \nabla \varphi_j \,:\, j = 1, ..., J, i = 1, ..., \infty\}$ is a **frame** for the L^2 space of vector fields on $\mathcal{M}$.
- A **frame** is an overcomplete spanning set commonly used in Harmonic analysis, must satisfy the frame inequalities:

$$A||v||^2 \leq \sum_{i,j} \left\langle v, \varphi_i \nabla \varphi_j \right\rangle^2 \leq B||v||^2$$

where $A, B > 0$ and $||\cdot||^2 = \langle \cdot, \cdot \rangle$ is the Hodge inner prod.

B & Giannakis, Spectral exterior calculus

The Spectral Exterior Calculus (SEC)

- We extend Thm to frames for Sobolev spaces of tensors
- SEC formulates the entire exterior calculus in these frames
- Key accomplishment: Representation of the 1-Laplacian

$$\Delta_1 = d\delta + \delta d$$

- Key challenge: Frame representations are not unique, requires Sobolev regularizations for numerical stability

B & Giannakis, Spectral exterior calculus

A CALCULUS NEEDS FORMULAS!

Object	Symbolic	Spectral
Function	f	$\hat{f}_k = \langle \phi_k, f \rangle_{L^2}$
Laplacian	Δf	$\langle \phi_k, \Delta f \rangle_{L^2} = \lambda_k \hat{f}_k$
L^2 Inner Product	$\langle f, h \rangle_{L^2}$	$\sum_i \hat{f}_i^* \hat{h}_i$
Dirichlet Energy	$\langle f, \Delta f \rangle_{L^2}$	$\sum_i \lambda_i \lvert\hat{f}_i\rvert^2$
Multiplication	$\phi_i \phi_j$	$c_{ijk} = \left\langle \phi_i \phi_j, \phi_k \right\rangle_{L^2}$
Function Product	fh	$\sum_{ij} c_{kij} \hat{f}_i \hat{h}_j$
Riemannian Metric	$\nabla \phi_i \cdot \nabla \phi_j$	$g_{kij} \equiv \left\langle \nabla \phi_i \cdot \nabla \phi_j, \phi_k \right\rangle_{L^2}$ $= \frac{1}{2}(\lambda_i + \lambda_j - \lambda_k) c_{kij}$
Gradient Field	$\nabla f(h) = \nabla f^* \cdot \nabla h$	$\langle \phi_k, \nabla f(h) \rangle_{L^2} = \sum_{ij} g_{kij} \hat{f}_i \hat{h}_j$
Exterior Derivative	$df(\nabla h) = df^* \cdot dh$	$\sum_{ij} g_{kij} \hat{f}_i \hat{h}_j$
Vector Field (basis)	$v(f) = v^* \cdot \nabla f$	$\sum_j v_{ij} \hat{f}_j$
Divergence	$\mathrm{div} v$	$\langle \phi_i, \mathrm{div} v \rangle_{L^2} = -v_{0i}$
Frame Elements	$b_{ij}(\phi_l) = \phi_i \nabla \phi_j(\phi_l)$	$G_{ijkl} \equiv \left\langle b_{ij}(\phi_l), \phi_k \right\rangle_{L^2} = \sum_m c_{mik} g_{mjl}$
Vector Field (frame)	$v(f) = \sum_{ij} v^{ij} b_{ij}(f)$	$\langle \phi_k, v(f) \rangle_{L^2} = \sum_{ijl} G_{ijkl} v^{ij} \hat{f}_l$
Frame Elements	$b^{ij}(v) = b^i \, db^j(v)$	$\left\langle \phi_k, b^{ij}(v) \right\rangle_{L^2} = \sum_{nlm} c_{kmi} G_{nlmj} v^{nl}$
1-Forms (frame)	$\omega = \sum_{ij} \omega_{ij} b^{ij}$	$\langle \phi_k, \omega(v) \rangle_{L^2} = \sum_{ij} \omega_{ij} \left\langle \phi_k, b^{ij}(v) \right\rangle_{L^2}$

Operator	Tensor	Symmetries
Quadruple Product	$c^0_{ijkl} = \left\langle \phi_i \phi_j, \phi_k \phi_l \right\rangle_{L^2} = \sum_s c_{ijs} c_{skl}$	Fully symmetric
Product Energy	$c^p_{ijkl} = \left\langle \Delta^p(\phi_i \phi_j), \phi_k \phi_l \right\rangle_{L^2} = \sum_s \lambda^p_s c_{ijs} c_{skl}$	(1,2), (3,4), (1,3), (2,4)
Hodge Grammian	$G_{ijkl} = \left\langle b^{ij}, b^{kl} \right\rangle_{L^2_1} = \frac{1}{2}\left[(\lambda_j + \lambda_l) c^0_{ijkl} - c^1_{ijkl}\right]$	(1,3), (2,4)
Antisymmetric	$\hat{G}_{ijkl} = \langle \hat{b}^{ij}, \hat{b}^{kl} \rangle_{L^2_1} = G_{ijkl} + G_{jilk} - G_{jikl} - G_{ijlk}$	(1,3), (2,4)
Dirichlet Energy $E_{ijkl} = \left\langle b^{ij}, \Delta_1(b^{kl}) \right\rangle_{L^2_1}$	$E_{ijkl} = \frac{1}{4}\left[(\lambda_i + \lambda_j + \lambda_k + \lambda_l)(c^1_{iljk} - c^1_{ikjl}) + (\lambda_j + \lambda_l - \lambda_i - \lambda_k) c^1_{ijkl} + (c^2_{ijkl} + c^2_{ikjl} - c^2_{iljk})\right]$	(1,3), (2,4)
Antisymmetric	$\hat{E}_{ijkl} = \langle \hat{b}^{ij}, \Delta_1 \hat{b}^{kl} \rangle_{L^2_1} = (\lambda_i + \lambda_j + \lambda_k + \lambda_l)(c^1_{iljk} - c^1_{ikjl}) + (c^2_{ikjl} - c^2_{iljk})$	(1,3), (2,4)
Sobolev H^1 Grammian	$G^1_{ijkl} = E_{ijkl} + G_{ijkl}, \quad \hat{G}^1_{ijkl} = \hat{E}_{ijkl} + \hat{G}_{ijkl}$	(1,3), (2,4)

Object	Symbolic	Spectral
Multiple Product	$c^0_I = \left\langle b^{i_0} \cdots b^{i_k}, 1 \right\rangle_H$	$c^0_I = \sum_s c_{i_0 i_1 s} c^0_{s i_2 \cdots i_k}$
Tensor	$H^{\tilde{I}\tilde{J}} = (db^{i_1} \cdot db^{j_1}) \cdots (db^{i_k} \cdot db^{j_k})$	$\hat{H}^{\tilde{I}\tilde{J}}_l \equiv \langle H^{\tilde{I}\tilde{J}}, b^l \rangle_H$
Evaluation	$= \nabla b^{i_1} \otimes \cdots \otimes \nabla b^{i_k} (b^{j_1}, \ldots, b^{j_k})$	$= \sum_{n=1}^{k^2} \prod_{s,r=1}^{k} g_{i_s j_r m_n} c_{l m_1 \cdots m_{k^2}}$
Tensor Product	$b_J = b^{j_0} \nabla b^{j_1} \otimes \cdots \otimes \nabla b^{j_k}$	$\left\langle b_J(b^{i_1}, \ldots, b^{i_k}), b^l \right\rangle = \sum_s \hat{H}^{\tilde{J}\tilde{I}}_s c_{s j_0 l}$
Frame Elements	$b^I = b^{i_0} db^{i_1} \wedge \cdots \wedge db^{i_k}$	$\left\langle b^I(b_J), b^l \right\rangle_H = \left\langle b^I \cdot b^J, b^l \right\rangle_H$
Riemannian Metric	$b^I \cdot b^J = b^{i_0} b^{j_0} \det([db^{i_q} \cdot db^{j_n}])$	$\left\langle b^I \cdot b^J, b_l \right\rangle_H = \sum_s \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) c_{s i_0 j_0 l} \hat{H}^{\tilde{I}\sigma(\tilde{J})}_s$
Hodge Grammian	$G_{IJ} = \left\langle b^I, b^J \right\rangle_{H_k} = \left\langle b^I \cdot b^J, 1 \right\rangle_H$	$\sum_s \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) c_{s i_0 j_0} \hat{H}^{\tilde{I}\sigma(\tilde{J})}_s$
d-Energy	$E^d_{IJ} = \left\langle db^I, db^J \right\rangle_{H_{k+1}}$	$\left\langle db^I \cdot db^J, 1 \right\rangle_{H_{k+1}} = \hat{H}^{IJ}_0$

BACK TO BASIS

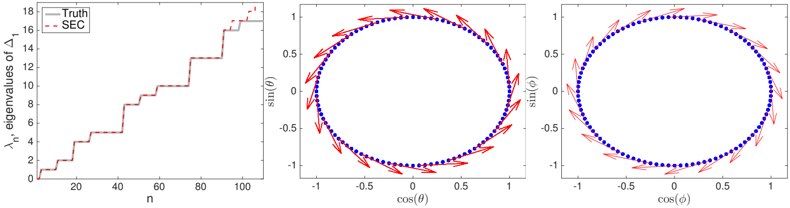
- We need the frame representation to build the 1-Laplacian

$$\Delta_1 = d\delta + \delta d$$

- Eigenfields of $\Delta_1 \Rightarrow$ smoothest basis for vector fields
- Can use to smooth vector fields and represent operators

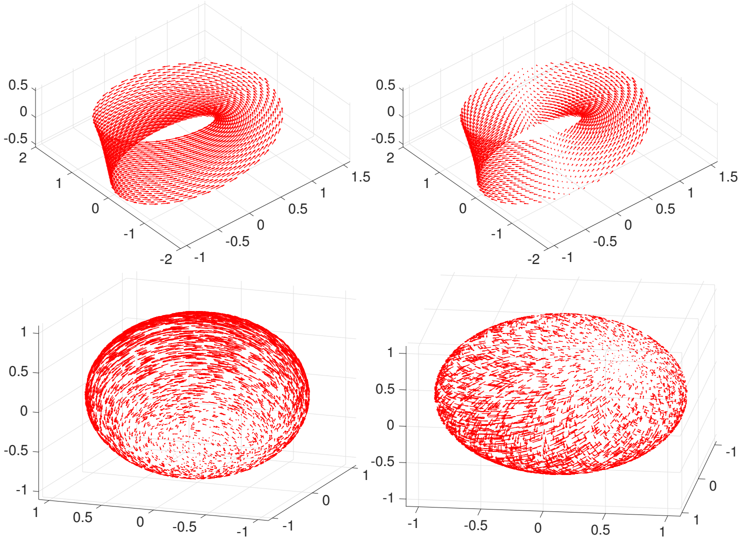
Numerical Verification on Flat Torus

Captures the true spectrum of the Hodge Laplacian.

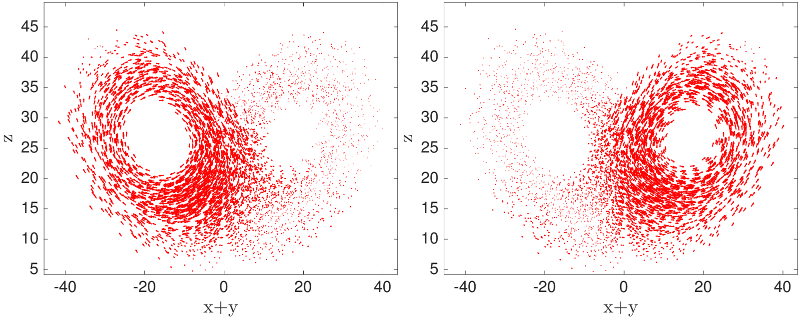


Harmonic forms correspond to unique homology classes.

Smoothest vector fields on the manifold



SEC is applicable to any data set



Matlab Code: http://math.gmu.edu/˜berry/

APPLYING THE SEC TO DYNAMICAL SYSTEMS

- Smooth/Denoise vector fields using SEC basis
- Next Step: Hodge decomposition

$$v = \nabla U + \delta A + v^{\perp}$$

- U is a potential, A is a tensor field, and $\Delta_1 v^{\perp} = 0$

Decomposing SDE Components

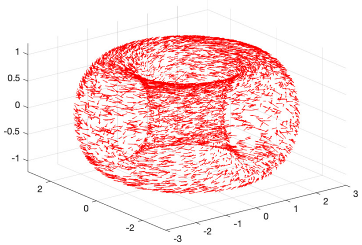
- Given a realization of an SDE on a manifold:

$$dx = f(x)\,dt + B(x)\,dW_t$$

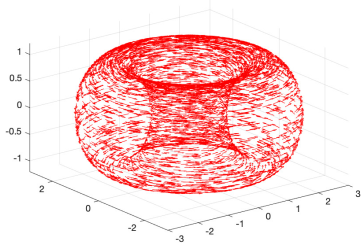
- Want to extract the deterministic component, $f(x)$
- Finite differences $x(t+\tau) - x(t) \approx f(x(t))$ but noisy
- Can smooth component functions using DM basis
- Better to smooth with SEC eigenvectorfields

Decomposing SDE Components

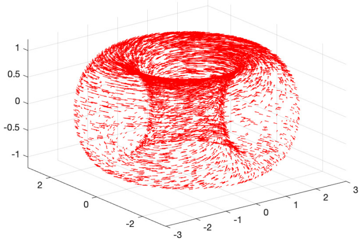
Finite Difference Est.



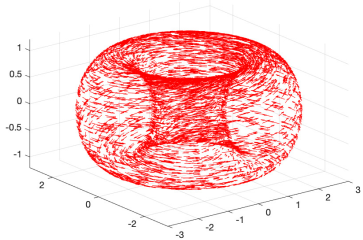
True Vector Field



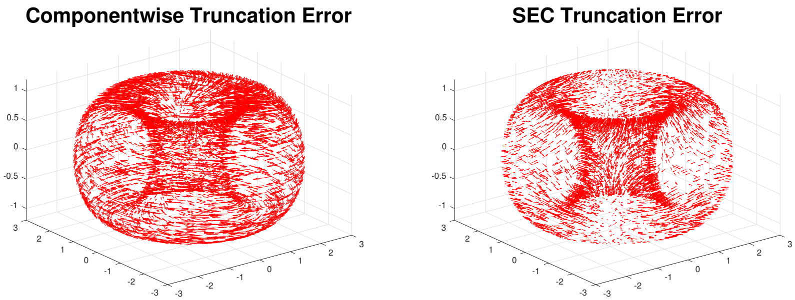
Componentwise Truncation



SEC Truncation



Decomposing SDE Components



SUMMARY

Manifold learning interweaves mathematics and statistics

- Kernel density estimation can be done on Riemannian manifolds with boundary
- Kernel matrices estimate geometric operators which contain all information about a manifold
- Statistics tell us when the algorithms will actually work
- Balloon estimators required for some operator estimation
- SEC builds the exterior calculus from the Laplacian

NEXT STEPS...

- Curse-of-intrinsic-dimensionality is still a problem
- We need better priors!
  - Smoothness $\Rightarrow$ Higher-order kernels
  - Symmetry $\Rightarrow$ Group/Lie structure
  - High curvature embeddings $\Rightarrow$ Projections?
  - Multi-scale structure $\Rightarrow$ Hierarchical kernels/Deep learning
- Beyond manifolds:
  - Sampling measures concentrated near manifolds
  - Metric-measure spaces, fractals, graphs