Geometry of Kernels

Tyrus Berry
George Mason University

October 14, 2015
Local Kernels

Geometry of Local Kernels
Diffeomorphism Representation
Example: Variable Bandwidth Kernels

Learning the Tangent Space

Kernel Weighted Vectors
Singular Value Scaling Laws

Iterated Diffusion Map

Rank-deficient Kernels
Geometric Flow for Non-Diffeomorphic Maps

Future Directions

Higher Order Kernels and Smoothness Priors
Higher Order Laplacians and Persistent Homology
Coarse Geometry of Density Scaling Laws
Local Kernels

A *local kernel* is a map $K : [0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that

$$K(\delta, x, x + \delta z) < ae^{-b||z||^2}$$

for some $a, b > 0$

More generally, we require the moments

$$m_\alpha(x) = \int_{z \in T_x \mathcal{M}} \left( \prod_{j=1}^m z_j^{\alpha_j} \right) K(\delta, x, x + \delta z) \, dz$$

to be finite

Prototypical Kernels: For $A(x) \in \mathbb{R}^{n \times n}$ and $b(x) \in \mathbb{R}^n$

$$K(\delta, x, y) = \exp \left( -\frac{1}{\delta^2} (y - x - \delta b(x))^\top A(x)^{-1} (y - x - \delta b(x)) \right)$$

We want analyze the diffusion maps algorithm for this more general class of kernels
Local Kernels

Prototypical Kernels: For $A(x) \in \mathbb{R}^{n \times n}$ and $b(x) \in \mathbb{R}^n$

$$K(\delta, x, y) = \exp \left( -\frac{1}{\delta^2} (y - x) \top A(x)^{-1} (y - x) \right)$$

Let $\{x_i\}$ be sampled from a compact manifold $\mathcal{M} \subset \mathbb{R}^n$

$\mathcal{M}$ has a geometry $g$ inherited from the ambient space

Diffusion maps built a graph Laplacian $L$ and found $L \rightarrow \Delta g$

For a local kernel, $L \rightarrow \Delta \hat{g}$ where $\hat{g}$ is a new geometry

In fact, every geometry on $\mathcal{M}$ is accessible via a local kernel
Local Kernels: New Geometry

- Prototypical Kernels: For $A(x) \in \mathbb{R}^{n \times n}$ and $b(x) \in \mathbb{R}^n$

  $$K(\delta, x, y) = \exp \left( -\frac{1}{\delta^2} (y - x) ^\top A(x)^{-1} (y - x) \right)$$

- The matrix $A$ will change the geometry

- $A(x)$ is an $n \times n$ matrix, but the metric $g_{ij}(x)$ is $m \times m$

- Need to restrict $A$ to the tangent space, let $\mathcal{I} : \mathbb{R}^n \rightarrow T_x M$

- We define $\hat{A}(x) = \mathcal{I}(x) A(x) \mathcal{I}(x)^\top$

- The new geometry is $\hat{g}(x) = A(x)^{-1/2} g(x) A(x)^{-1/2}$
Local Kernels: Idea of Proof

- Local property allows localization, in Taylor expansion we find,

\[
\int_{T_x \mathcal{M}} K(\delta, x, x+\delta u) \left( f(x) + u^T \nabla f(x) + \sum_{i,j} u_i u_j \frac{\partial^2 \tilde{f}(0)}{\partial s_i \partial s_j} \right) du
\]

- Moving the integral in we find

\[
(K \tilde{f})_i \propto m_0(x_i)f(x_i) + \delta^2 m_0(x_i) \left( \sum_{ij} \tilde{A}_{ij} \frac{\partial^2 \tilde{f}(0)}{\partial s_i \partial s_j} + \omega(x_i)f(x_i) \right)
\]

- Next we analyze the operator \( \sum_{ij} \tilde{A}_{ij} \frac{\partial^2 \tilde{f}(0)}{\partial s_i \partial s_j} \)
Local Kernels: Idea of Proof

- Next we analyze the operator $\sum_{ij} \hat{A}_{ij} \frac{\partial^2 \tilde{f}(0)}{\partial s_i \partial s_j}$
- By changing variables $\hat{s}_i = \sum_j (\hat{A}^{-1/2})_{ij} s_j$ we find

$$\sum_{ij} \hat{A}_{ij} \frac{\partial^2 \tilde{f}}{\partial s_i \partial s_j} = \sum_i \frac{\partial^2 \tilde{f}(0)}{\partial \hat{s}_i^2} + \kappa \cdot \nabla f = \Delta \hat{g} f + \kappa \cdot \nabla f$$

- The new coordinates $\hat{s}_i$ are the geodesic coordinates of $\hat{g}$
- We can rewrite $\kappa \cdot \nabla f = \tilde{\omega}(x)f(x) + m_0 \cdot \nabla f$
- $m_0(x)$ is the change of volume form $d\hat{V}(x) = m_0(x)dV(x)$
- $m_0 \cdot \nabla f$ cancelled by right normalization (sampling $d\hat{V}(x)$)
- $\tilde{\omega}(x)f(x)$ cancelled by left normalization
Assume we have two data sets \( \{ x_i \}_{i=1}^N \) and \( \{ \tilde{x}_i \}_{i=1}^N \) which are related by an unknown diffeomorphism \( x_i = \mathcal{H}(\tilde{x}_i) \).

First assume that \( \mathcal{H} \) is an isometry.

Diffusion maps algorithms we finds \( L \rightarrow \Delta \) and \( \tilde{L} \rightarrow \tilde{\Delta} \).

Since \( \mathcal{H} \) is an isometry \( \tilde{g} = g \) so \( \Delta f(x) = \tilde{\Delta}(f \circ \mathcal{H}) = \tilde{\Delta}(\tilde{f}(\tilde{x})) \).

The eigenfunctions \( \varphi_i(x) \) and \( \tilde{\varphi}_i(\tilde{x}) \) are the same up to orthogonal transformation for repeated eigenvalues.
Assume we have two data sets \( \{ x_i \}_{i=1}^N \) and \( \{ \tilde{x}_i \}_{i=1}^N \) which are related by an unknown diffeomorphism \( x_i = \mathcal{H}(\tilde{x}_i) \).

First assume that \( \mathcal{H} \) is an isometry, we have:

\[
\begin{align*}
\mathcal{M} \xrightarrow{\mathcal{H}} \mathcal{H}(\mathcal{M}) \\
\downarrow \tilde{\Phi} \quad \quad \quad \quad \quad \quad \downarrow \Phi \\
L^2(\mathcal{M}, \tilde{g}) \approx \mathbb{R}^{\hat{n}} \xrightarrow{U} L^2(\mathcal{H}(\mathcal{M}), g) \approx \mathbb{R}^{\hat{m}}
\end{align*}
\]

- \( \tilde{\Phi} \) and \( \Phi \) are diffusion maps with \( t = 0 \)
- \( U \) is an orthogonal linear transformation, easy to fit
Diffeomorphism Representation

- Assume we have two data sets \( \{x_i\}_{i=1}^N \) and \( \{\tilde{x}_i\}_{i=1}^N \) which are related by an unknown diffeomorphism \( x_i = \mathcal{H}(\tilde{x}_i) \).
- We want to learn \( \mathcal{H} \), we find \( D\mathcal{H}(x_i) \) by a kernel weighted linear regression (more on this later).
- Use a local kernel to pullback the geometry of \( \mathcal{H}(\mathcal{M}) \) onto \( \mathcal{M} \).
- We want \( \tilde{g}(u, v) = g(D\mathcal{H}u, D\mathcal{H}v) \) so that \( \tilde{g} = D\mathcal{H}^\top gD\mathcal{H} \).
- With this new geometry, \( \mathcal{M} \) and \( \mathcal{H}(\mathcal{M}) \) are isometric!
- Use standard diffusion maps on \( \mathcal{H}(\mathcal{M}) \) to get \( \Delta_g \).
- Local kernel on \( \mathcal{M} \) with \( \hat{A}(x) = D\mathcal{H}(x)^\top D\mathcal{H}(x) \) yeilds \( \Delta_{\tilde{g}} \).
- Find the linear map between eigenfunctions of \( \Delta_g \) and \( \Delta_{\tilde{g}} \).
Diffeomorphism Representation

Diffusion Map + Linear Map

Diffusion Map

Local Kernel + Linear Map
Variable bandwidth kernels are local kernels with diagonal covariance,

\[ K^{S}_{\epsilon}(x, y) = h \left( \frac{\|x - y\|^2}{\epsilon q_\delta(x)^{\beta} q_\delta(y)^{\beta}} \right) \]

- Diagonal covariance is *conformal* (only changes the volume)
- VB uses different right normalization to find \( \Delta_g \) instead of \( \Delta_{\hat{g}} \)
- Changes inside vs. outside the exponential can cancel out in the asymptotics but still effect variance!
- Need to analyze the bias/Chernoff bounds for local kernels
- This is what prevents application to non-compact manifolds
Derivative Estimation

- Let’s revisit any smooth mapping $y_i = \mathcal{H}(x_i)$, let

$$D(x) = \sum_{i=1}^{N} \exp \left( -\frac{||x_i - x||^2}{2\epsilon} \right).$$

- Let $X$ to be a matrix with columns

$$X_j = D(x)^{-1/2} \exp \left( -\frac{||x_j - x||^2}{4\epsilon} \right) (x_j - x)$$

- Let $Y$ be a matrix with columns

$$Y_j = D(x)^{-1/2} \exp \left( -\frac{||x_j - x||^2}{4\epsilon} \right) (y_j - y)$$

- Then $\lim_{N \to \infty} \frac{1}{\epsilon} YX^\top = D\hat{\mathcal{H}}(x) + \epsilon R_\mathcal{H}(x) + O(\epsilon^2)$
Tangent Space Estimation

- When \( y_i = x_i \), we have
  \[
  \lim_{N \to \infty} \frac{1}{\epsilon} XX^T = I(x)^T I(x) + \epsilon R_I(x) + O(\epsilon^2)
  \]

- If \( v \in T_{x.\mathcal{M}} \), then \( v^T XX^T v = \epsilon \|v\|^2 + O(\epsilon^2) \),
- If \( v \in T_{x.\mathcal{M}^\perp} \) we find \( v^T XX^T v = O(\epsilon^2 \|v\|^2) \).
- If \( v \) is a singular vector, the associated singular value,
  \[
  \sigma_v = \lim_{N \to \infty} \frac{\sqrt{v^T XX^T v}}{\|v\|},
  \]
  is order-\( \sqrt{\epsilon} \) if \( v \in T_{x.\mathcal{M}} \), or order-\( \epsilon \) if \( v \in T_{x.\mathcal{M}^\perp} \).
Local Kernels

Learning the Tangent Space

Iterated Diffusion Map

Future Directions

Kernel Weighted Vectors

Singular Value Scaling Laws

Tyrus Berry  George Mason University

Geometry of Kernels
Feature Identification

- Extend local kernel representation to rank-deficient maps
- Idea: \( y_i = \mathcal{H}(x_i) \) is a lower-dimensional feature of interest
- Learn the part of the manifold \( \mathcal{M} \ni x_i \) which determines \( y_i \)
- If \( y_i \) is invariant under some transformations of \( \mathcal{M} \) we want to “mod out” by these transformations
- Problem is that \( \mathcal{H} \) is no longer a diffeomorphism (not even same dimension!)
- We can still estimate \( D\mathcal{H} \) but no pullback geometry
Rank-deficient Kernels

- We can still estimate $D\mathcal{H}$ but no pullback geometry
- Problem is that $D\mathcal{H}^\top D\mathcal{H}$ does not define a local kernel
- Not full rank: kernel does not decay in all directions
- Instead we regularize the covariance:

$$C_{\mathcal{H}(0)}(x) = \left((1 - \tau)I_{m \times m} + \tau D\hat{\mathcal{H}}(x)^\top D\hat{\mathcal{H}}(x)\right)^{-1}$$

- Small perturbation of identity, emphasizes direction of importance and $C_{\mathcal{H}(0)}(x)$ defines a local kernel
Instead we regularize the covariance:

\[ C_{\mathcal{H}(0)}(x) = \left( (1 - \tau)I_{m \times m} + \tau D\hat{\mathcal{H}}(x)\mathbf{\top} D\hat{\mathcal{H}}(x) \right)^{-1} \]

New geometry is \( g_{\mathcal{H}(0)} = c_{\mathcal{H}(0)}^{-1/2} g_{\mathcal{H}(0)} c_{\mathcal{H}(0)}^{-1/2} \) where

\[ c_{\mathcal{H}(0)}^{-1/2} = ((1 - \tau)I_{d \times d} + \tau D\mathcal{H} \mathbf{\top} D\mathcal{H})^{1/2} \]

New embedding with respect to this embedding

\[ \Phi_{s}^{(0)}(x) = (e^{s\lambda_1} \varphi_1(x), ..., e^{s\lambda_M} \varphi_M(x)) \mathbf{\top} \equiv x^{(1)} \]

Still have a map \( y_i = \mathcal{H}^{(1)}(x_i^{(1)}) = \mathcal{H}(\Phi^{(0)}(x_i)) \), so repeat!
Still have a map $y_i = \mathcal{H}^{(1)}(x_i^{(1)}) = \mathcal{H}(\Phi^{(0)}(x_i))$, so repeat!

At each step we have a local kernel which increases the emphasis on directions of interest:

We have $g(t + \tau) = c(x, t + \tau)^{-1/2} g(t) c(x, t + \tau)^{-1/2}$ so

$$g(t + \tau) = g(t) + \frac{\tau}{2} \left( D\mathcal{H}^\top D\mathcal{H} g(t) + g(t) D\mathcal{H}^\top D\mathcal{H} - 2g(t) \right) + \mathcal{O}(\tau^2)$$

We interpret this Iterated Diffusion Map as a discretized geometric flow

$$\frac{dg}{dt} = \lim_{\tau \to 0} \frac{g(t + \tau) - g(t)}{\tau} = -g + \frac{1}{2} \left( D\mathcal{H}^\top D\mathcal{H} g + g D\mathcal{H}^\top D\mathcal{H} \right)$$
Assume $\mathcal{M} = \mathcal{N} \times \mathcal{P}$, where $\mathcal{N}$ is the feature space and $\mathcal{P}$ are irrelevant

\[
\dot{g} = \frac{1}{2} \left( (D\mathcal{H}^\top D\mathcal{H} - I)g + g(D\mathcal{H}^\top D\mathcal{H} - I) \right)
\]

\[
= \frac{1}{2} \left( \begin{pmatrix} 0 & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} g_\mathcal{N} & 0 \\ 0 & g_\mathcal{P} \end{pmatrix} + \begin{pmatrix} g_\mathcal{N} & 0 \\ 0 & g_\mathcal{P} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -I \end{pmatrix} \right),
\]

This implies that $\dot{g}_\mathcal{N} = 0$ and $\dot{g}_\mathcal{P} = -g_\mathcal{P}$
For product manifolds we can recover the local kernels diagram using the Iterated Diffusion Map on the product space!

\[ M = \mathcal{N} \times \mathcal{P} \quad \xrightarrow{\mathcal{H}} \quad \mathcal{N} \]

\[ \psi \equiv \lim_{\ell \to \infty, s \to 0} \Phi_{s}^{(\ell)} \circ \cdots \circ \Phi_{s}^{(0)} \quad \xrightarrow{\tilde{\Phi}} \quad \tilde{\Phi} \]

\[ L^{2}(\mathcal{N}, \tilde{g}) \approx \mathbb{R}^{M} \quad \xrightarrow{H} \quad L^{2}(\mathcal{N}, g_{\mathcal{N}}) \approx \mathbb{R}^{M} \]
Geometric Flow on the Annulus
Geometric Flow on the Annulus
Geometric Flow on the Torus
Geometric Flow on the Torus
Geometric Flow on the Sphere
Geometric Flow on the Sphere
Higher Order Kernels

- Higher order kernels have more zero moments
  \[(y - x)e^{-||y-x||^2}\]
- For density estimation, an order \(p\) kernel has bias \(O(\delta^p)\)
- Assumes additional smoothness on the density
- Should be able to do same thing, assuming smooth manifold
- Curse-of-dimensionality: error \(\approx O(\delta^p, N^{-1}\delta^{-d})\)
- Take \(p = d\) and \(\delta = O(N^{-1/(2d)})\), error becomes \(O(N^{-1/2})\)
Higher Order Laplacians

- It is possible to define higher order Laplacians $\Delta_k$ which act on differential $k$-forms.

- The kernel of these Laplacians are isomorphic to the $k$-th de Rahm cohomology.

- Technically, the zero Laplacian contains all geometric information.

- Unclear how to construct higher order Laplacians from lower order.

- Alternative construction in terms of Discrete Exterior Calculus.

- Kernel based Homology would allow persistence in terms of $\delta$. 
Coarse Geometry

- Differential is a nice starter home, but real data doesn’t really live there.
- Real data lives on the support of a measure.
- Measure may be ‘concentrated near’ a manifold.
- Exact tangent spaces yield certain scaling laws.
- Can define more general tangent space, allow any scaling law.
- Some kind of generalization of a manifold?