Convergence of Graph Laplacians

Daniel Ting [1]

joint work with
Ling Huang [2] and Michael Jordan [3]

[1] Dept. of Statistics, UC Berkeley

[2] Intel Research, Berkeley

[3] EECS and Dept. of Statistics, UC Berkeley

July 29, 2011

Overview

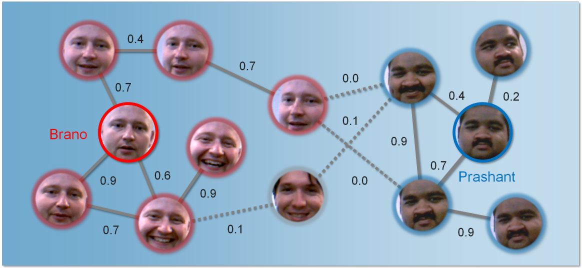
What the talk is about

- My general goal: Understand methods that rely on the manifold assumption

Graph Laplacians

- Used in:
  - Clustering
  - Semi-supervised learning
  - Non-linear Dimensionality Reduction
- Goals:
  - Better understanding of Laplacian or “Laplacian like” methods
  - Analysis of kNN graphs
  - Better choices when constructing graphs

Graph Laplacian



Definition (Graph Laplacian)

$$W = \text{edge weight matrix}$$

$$D = \text{diagonal degree matrix}$$

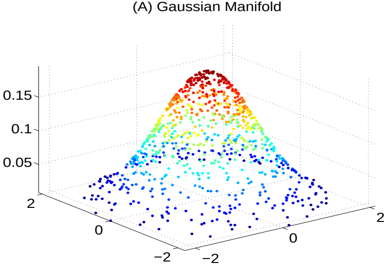
There are three commonly used versions of Graph Laplacians

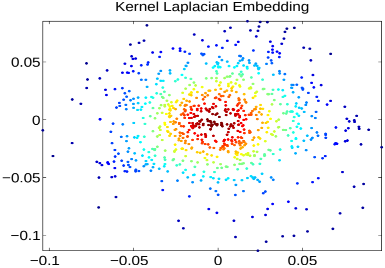
$$L_u = D - W \quad \text{(unnormalized)}$$

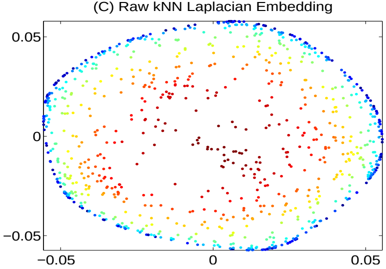
$$L_n = I - D^{-1/2} W D^{-1/2} \quad \text{(normalized)}$$

$$L_{rw} = I - D^{-1} W \quad \text{(random walk).}$$

A visual example: kNN non-linear embedding







Two graph constructions:
1) Why do they behave differently?
2) Is it fixable?

Constructing graphs on a manifold

Convert Euclidean distance to edge weights: Existing theory

1. Choose a smooth kernel K (e.g. Gaussian)
2. Choose a bandwidth h
3. Choose a weighting/normalization exponent $1/d(x)^\alpha$

$$w_{ij} = \frac{1}{d(x_i)^\alpha d(x_j)^\alpha} K\left(\frac{\|x_i - x_j\|}{h}\right)$$

Overly restrictive conditions!

- Fails to cover kNN graphs
- Important since kNN graphs
  - are sparse and have good computational properties
  - show better empirical performance in SSL applications

Constructing graphs on a Manifold

Generalizing edge weight choices

1. Allow for a non-smooth kernel K
2. Use a location dependent bandwidth hR
3. Use arbitrary weight functions ω

$$w_{ij} = \omega(x_i)\omega(x_j)\ K\left(\frac{\|x_i - x_j\|}{hR(x_i, x_j)}\right)$$

Covers most geometric graph constructions of interest. This covers kNN graphs by using an indicator kernel.

Overview: The details

Analysis

- Introduce kernel-free framework for analysis using diffusion theory
- Identify the limiting operator and corresponding smoothness functional
  - Effect of the density
  - Effect of the graph construction method
- Discuss how choose a good graph construction

Application

- kNN graphs
- "self-tuning" graphs of Zelnik-Manor & Perona (2004)
- Locally Linear Embedding (LLE)

Setting and Notation

Setting

- Compact, smooth m-dimensional manifold without boundary $\mathcal{M}$ embedded in $\mathbb{R}^k$
- Density p
- Graph construction method (e.g. kNN) where neighborhood sizes are shrinking
- Examine $L_u^{(n)} f \rightarrow ?$ (i.e. pointwise convergence in the strong operator topology)

Notation

Data points	$x, y \in \mathbb{R}^k$
Smooth C^3 function on $\mathcal{M}$	$f : \mathcal{M} \rightarrow \mathbb{R}$
Normal coordinates for y in neighborhood of x	$s \in \mathbb{R}^m$
Canonical measure on $\mathcal{M}$	η
Base kernel function	$K(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$
degree function	$d_n(\cdot)$

Motivation for a Kernel-free framework

Previous question

- Here's a graph construction method. Does a limit exist and what is it?

Broader questions

- What characterizes the limit?
- What are sufficient conditions for a limit to exist?

Answer: Diffusion processes

Characterizations

- Laplacian is a (negative) infinitesimal generator for a random walk
- Drift μ and diffusion $\sigma\sigma^T$ terms characterize a diffusion process

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

- Infinitesimal generator of a diffusion process is

$$\mathcal{G}f = \frac{1}{2}\sum_{i,j}\left(\sigma\sigma^T\right)_{ij}\frac{\partial^2 f}{\partial s_i \partial s_j} + \sum_i \mu_i \frac{\partial f}{\partial s_i}.$$

Interpreting the drift and diffusion

A special elliptic operator: Laplace-Beltrami

If $\sigma\sigma^T = g(\cdot)\mathbb{I}$ and μ/g is a conservative vector field with $\mu/g = 2\nabla \log q$, then

$$\mathcal{G} = \frac{1}{2} g\Delta + \langle \mu, \nabla \rangle = \frac{1}{2} g\Delta_q$$

Continuous analog to graph Laplacian

- Smoothness functional

$$\int \langle f, -\Delta_q f \rangle d\eta = \int \langle \nabla f, \nabla f \rangle q d\eta = \|\nabla f\|^2_{L_2(q)}$$

Main result

Key Assumptions

- Kernel K of bounded variation
- Bandwidth scaling $h_n \to 0$ and $h_n^{m+2} n / \log n \to \infty$
- Weight $\omega_n(x)$ and bandwidth functions $R_n(x, y)$ with Taylor-like expansions

$$R_n(x, x+\epsilon) = h_n\big(\mathbf{r}(\mathbf{x}) + \epsilon\, \mathbf{d_r}(\mathbf{x}, \mathbf{sign}(\mathbf{u}(\mathbf{x})^{\mathbf{T}}\epsilon)) + o(h_n)\big)$$

Result

Expressed in normal coordinates, the drift and diffusion defined by L_{rw} are

$$\mu_s(x) = r(x)^2 \left(\frac{\nabla p(x)}{p(x)} + \frac{\nabla \omega(x)}{\omega(x)} + \frac{m+2}{2} \frac{\overline{d_r}(x)}{r(x)} \right),$$

$$\sigma_s(x)\sigma_s(x)^T = r(x)^2 \mathbb{I}$$

where $\overline{d_r}(x) = \frac{1}{2}(d_r(x,1) + d_r(x,-1))u_s(x)$.

Remarks

Alternative form (for self-adjoint Laplacians)

If $R_n(x,y)/h_n = \sqrt{r(x)r(y)} + o(h_n)$ e.v. then we have

- The asymptotic limit of the graph Laplacian is

$$-c_n L_{rw} \to r^2 \Delta_q$$
$$-c_n' L_u \to \frac{q}{p} \Delta_q$$

  where $q = p^2 \omega^2 r^{m+2}$.
- Gives a *change of measure* from p to q

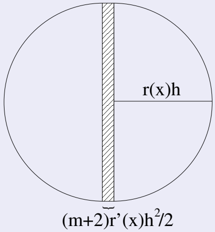
Proof sketch

- Based on one cute trick. Write a kernel with bounded variation as a weighted sum of indicator kernels

$$K(x) = \int \mathbb{I}(x < z) d\nu_+(z) - \int \mathbb{I}(x < z) d\nu_-(z)$$

- All the calculations involving non-random quantities reduce to finding moments for a sphere

- Location dependent bandwidth $\Longrightarrow$ indicator kernel behaves like it is shifted by $\frac{m+2}{2} h^2 d_r$. Shift introduces drift. Effect on the diffusion term is of smaller order than the leading term.



- Use Bernstein inequality and Borel-Cantelli to get almost sure convergence of random quantities

Application to kernel graphs

Weights

$$w_{ij} = \frac{K\left(\|x_i - x_j\| / h_n\right)}{d_n(x_i)^\alpha d_n(x_j)^\alpha}$$

$$\omega_n(x) = 1/d_n(x)^\alpha \to 1/p(x)^\alpha$$

Asymptotics (known)

- $R_n(x, y) = 1$
- $q = p^2 w^2 r^{d+2} = p^{2-2\alpha}$
- Asymptotic random walk and unnormalized Laplacians are

$$c_n L_{rw} \to \Delta_{p^{2-2\alpha}}$$

$$c'_n L_u \to p^{1-2\alpha} \Delta_{p^{2-2\alpha}}$$

- If $\alpha < 1$, drift *towards* high density regions.

Application to kNN

Weights

$$w_{ij} = \mathbb{I}\big(\, \|x_i - x_j\| < R_n(x_i, x_j)\big)$$
$$k_n/n \to 0 \quad \text{and} \quad k_n^{1+2/m}/(n^{2/m}\log n) \to \infty$$

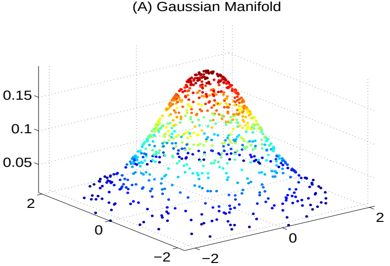
Undirected kNN graph (OR rule)

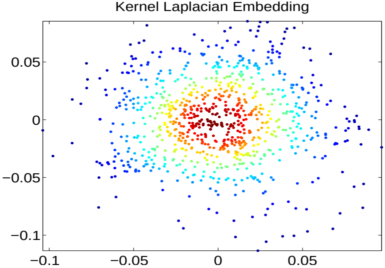
- $R_n(x, y)/h_n \approx \max\{p(x)^{-1/m}, p(y)^{-1/m}\}$
- $\overline{d_r}(x) = \frac{1}{2}\nabla p(x)^{-1/m}$
- Asymptotic random walk and unnormalized Laplacians are

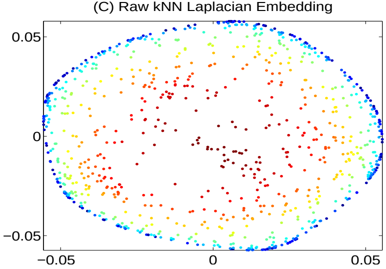
$$c_n L_{rw} = c'_n L_u \to \frac{1}{p^{2/m}} \Delta_{p^{1-2/m}}$$

- Drift *away from* high density regions if $m = 1$!

kNN non-linear embedding example





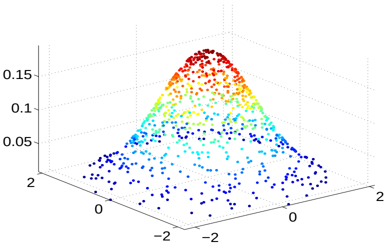


Fix:

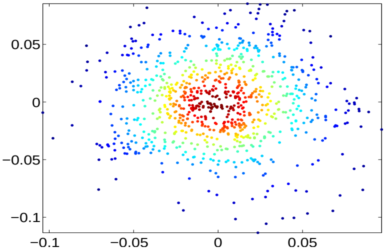
- kNN limit: $\frac{1}{p}\Delta_{p^0}$
- kernel limit: Δ_p
- Take $\omega = p^{1/2} \approx r^{-1}$

kNN non-linear embedding example

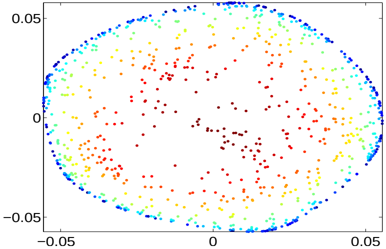
(A) Gaussian Manifold



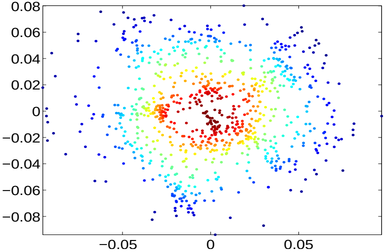
Kernel Laplacian Embedding



(C) Raw kNN Laplacian Embedding



(D) Rescaled kNN Laplacian Embedding



Application to "self-tuning" graphs

"self-tuning" kernel (Zelnik-Manor & Perona (2004))

$$K(x,y) = exp\left(-\frac{\|x-y\|^2}{\sigma_x \sigma_y}\right)$$

where σ_x is the distance between x and the k^{th} neighbor

Asymptotics

- $r(x,y) = \sqrt{p(x)^{-1/m} p(y)^{-1/m}}$
- $\overline{d_r}(x) = \frac{1}{2} \nabla p(x)^{-1/m}$ (same as undirected kNN)
- Same asymptotic Laplace-Beltrami operator as for kNN

$$c_n L_{rw} = c_n' L_u \rightarrow \frac{1}{p^{1+2/m}} \Delta_{p^{1-2/m}}$$

Application of kernel-free approach to LLE

Reminder: LLE

Goal: Find weights and coordinates that minimize reconstruction error.

Find weights (local regression)	Find coordinates (eigen-decomposition)
$\min_W Y^T(I-W)^T(I-W)Y$	$\min_z z^T(I-W)^T(I-W)z$

Behavior

Belkin & Niyogi (2003) give a heuristic derivation that LLE should behave like the Laplace-Beltrami operator, but... This is not completely true!

Application of kernel-free approach to LLE

Normal coordinates

- Converting normal coordinates s to extrinsic coordinates y
$$y = x + H_{T_x} s + L_{T_x^\perp}(ss^T) + O(\|s\|^3)$$

Rough analysis

- Examine the “drift” and “diffusion” terms for the LLE matrix $I - W$.
$$\mu_{\mathbb{R}^k}(x) = H_{T_x} \underbrace{\mathrm{E}s}_{\mu(x)} + L_{T_x^\perp} \underbrace{\mathrm{E}(ss^T)}_{\sigma\sigma^T(x)}$$
- The quantities in normal coordinates satisfy
$$\mu = 0 \qquad L_{T_x^\perp}\left(\sigma\sigma^T\right) = 0$$

Application of kernel-free approach to LLE

Implications for behavior

- $\mu = 0$
  - No effect of density on drift
- $L_{T_x^\perp}\left(\sigma\sigma^T\right) = 0$
  - Curvature of manifold affects diffusion term
  - No well-defined limit when L is not full rank
  - Does not behave like any elliptic operator if L is full rank

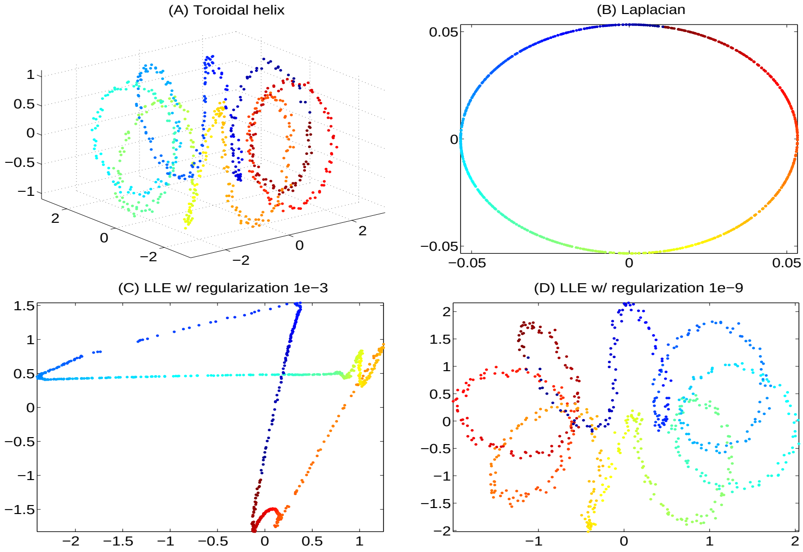
Behavior in practice

- Weights are regularized

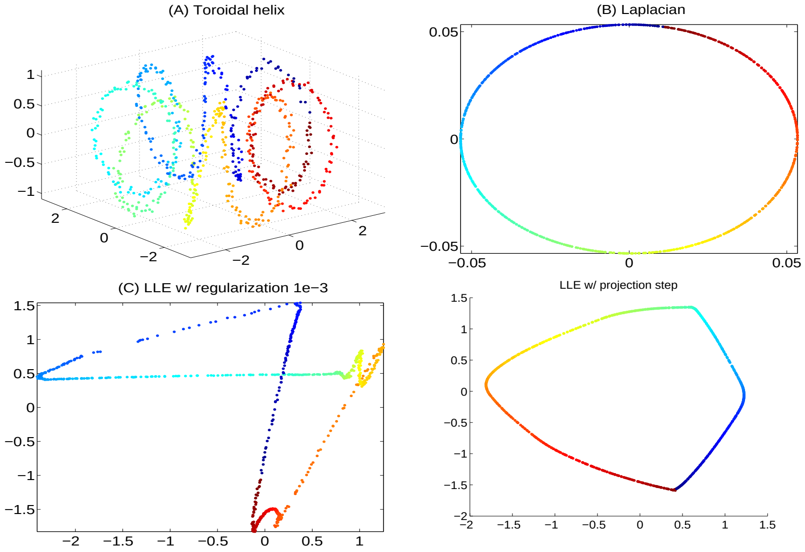
  $\Longrightarrow$ favors constant weights

  $\Longrightarrow$ like kNN-Laplacian if regularization is large

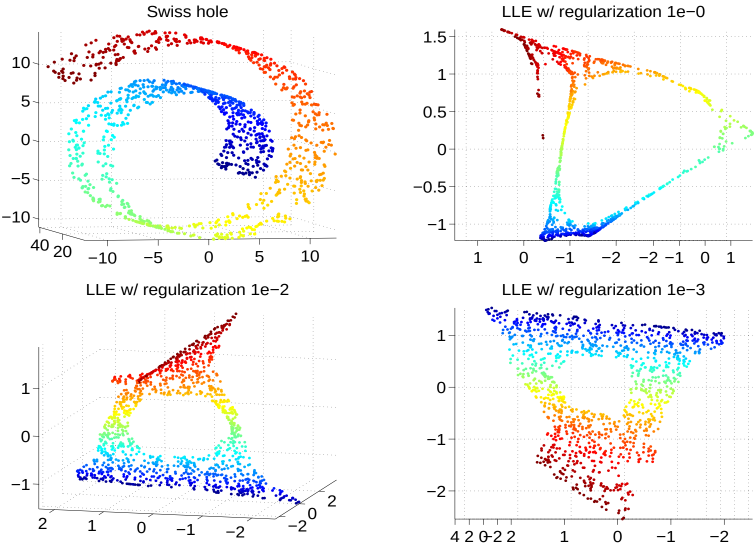
Explaining weird behavior of LLE



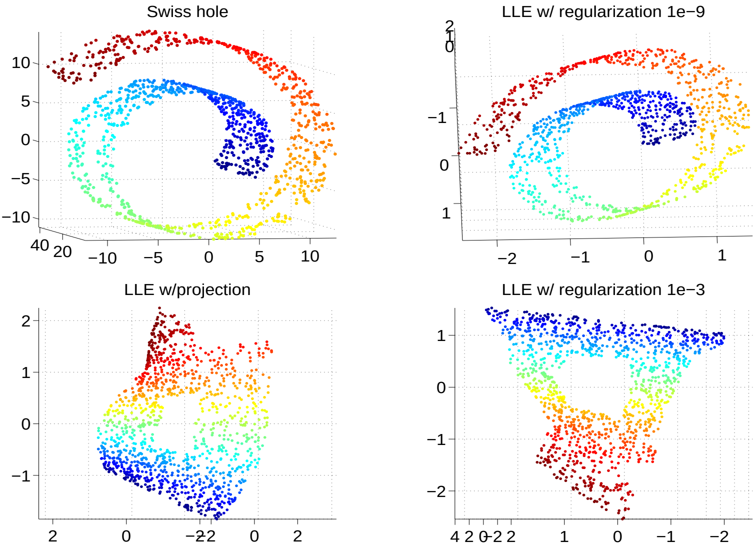
Explaining weird behavior of LLE



Effect of regularization and boundary in LLE



Effect of regularization and boundary in LLE



Other good properties: Convergence of Eigenvectors

Importance

- Construction of valid basis functions
- Possibly of relevance for spectral clustering

Graph Laplacians and spectrum (von Luxburg et al. (2008))

$$L_u f(\cdot) = d(\cdot)f(\cdot) - \int K(\cdot, y)f(y)dP_y$$

- No convergence for eigenvalues in $rng(d)$.
- Good spectral properties: Choose d so $rng(d) = \{1\}$.

Graph construction and spectrum

Asymptotic degree operator and weighting density

$$d(x) = p(x)\omega(x)^2 r(x)^m \overset{set}{=} 1$$
$$q(x) = p(x)^2\omega(x)^2 r(x)^{m+2}$$
$$= p(x)r(x)^2$$ Any density q is possible

Possible "good" symmetric graph constructions

	w/o loc dep bw	w/ loc dep bw
Unnormalized	Δ_p	$\frac{q}{p}\Delta_q$
Normalized	$p^{\frac{1}{2}-\alpha}\Delta_{p^{2-2\alpha}}p^{-\frac{1}{2}+\alpha}$	$g\left(\frac{q}{p}\Delta_q\right)g^{-1}$

Conclusion

What we introduce

- Kernel-free framework using diffusion processes
- Analysis for a general class of graph constructions
- Location dependent bandwidth + arbitrary weights + non-smooth kernels

Practical implications

- Better understanding of kNN graphs
- LLE
  - no “drift” component
  - affected by curvature of the manifold
- Construction of arbitrary first order smoothness functionals $\|\nabla f\|^2_{L_2(q)}$, not just $q = p^\alpha$.
- Pilot density estimates lead to “better” graph constructions

Belkin, M. and Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

Donoho, D.L. and Grimes, C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 100(10):5591, 2003.

von Luxburg, U., Belkin, M., and Bousquet, O. Consistency of spectral clustering. *Annals of Statistics*, 36(2):555–586, 2008.

Zelnik-Manor, L. and Perona, P. Self-tuning spectral clustering. In *NIPS 17*, 2004.

Zhang, Z. and Zha, H. Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment. *SIAM Journal on Scientific Computing*, 26:313, 2004.