An Introduction to Manifold Methods

Partha Niyogi
Mikhail Belkin
The University of Chicago

Collaborators: V. Sindhwani, X. He, D. Surendran, A. Jansen, S. Smale, S. Weinberger

High Dimensional Data

- Raw Format of Natural Data is often high dimensional.
- Curse of Dimensionality.
- Search for low dimensional structure and models.

Principal Components Analysis

Given $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^D$
Find $y_1, \ldots, y_n \in \mathbb{R}$ such that

$$y_i = \mathbf{w} \cdot \mathbf{x}_i$$

and

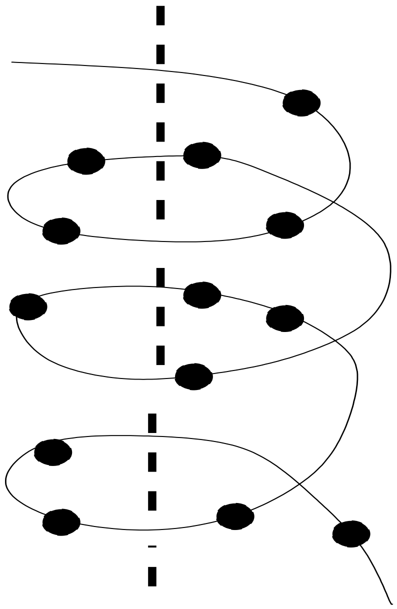
$$\max_{\mathbf{w}} \text{Variance}(\{y_i\}) = \sum_i y_i^2 = \mathbf{w}^T \left(\sum_i \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{w}$$

$$\mathbf{w}_* = \text{leading eigenvector of } \sum_i x_i x_i^T$$

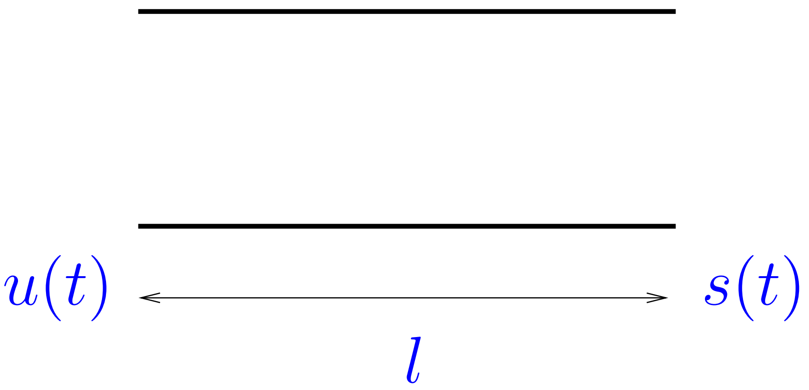
Manifold Model

Suppose data does not lie on a linear subspace.

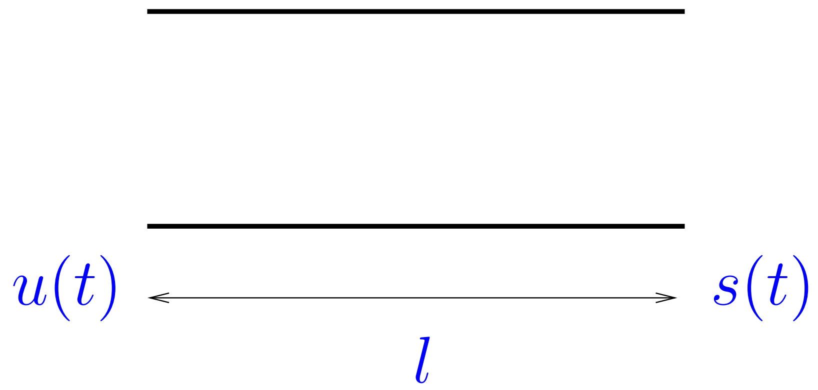
Yet data has inherently one degree of freedom.



An Acoustic Example



An Acoustic Example



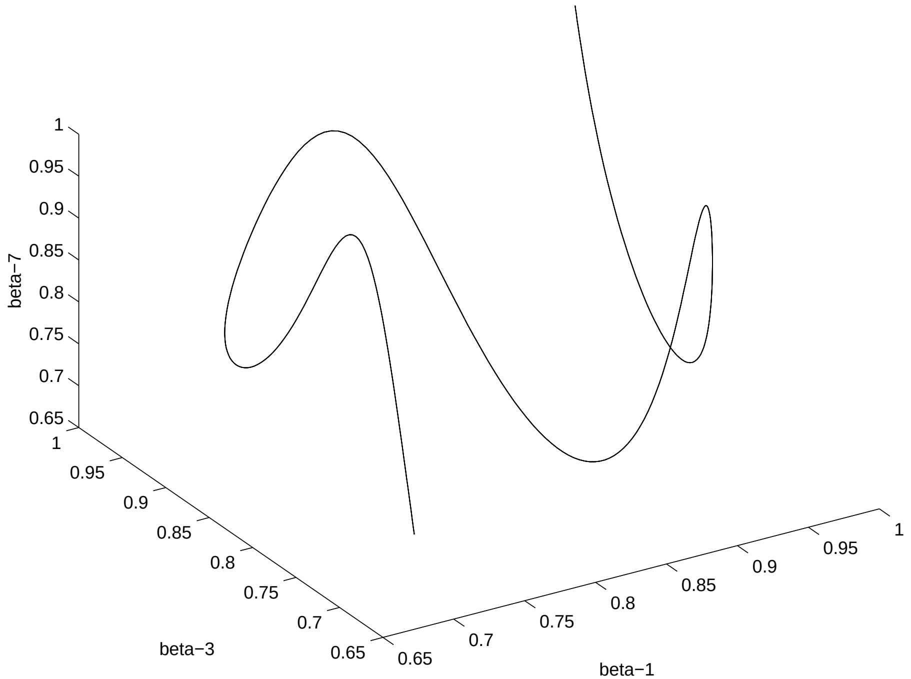
One Dimensional Air Flow

(i) $\frac{\partial V}{\partial x} = -\frac{A}{\rho c^2}\frac{\partial P}{\partial t}$

(ii) $\frac{\partial P}{\partial x} = -\frac{\rho}{A}\frac{\partial V}{\partial t}$

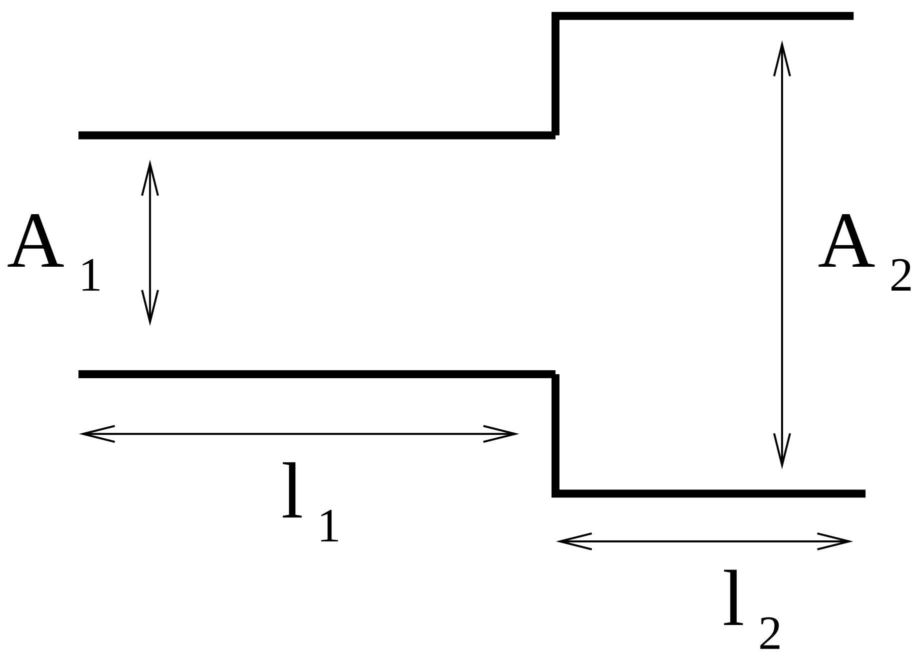
$V(x,t) =$ volume velocity
$P(x,t) =$ pressure

Solutions



$$u(t) = \sum_{n=1}^{\infty} \alpha_n \sin(n\omega_0 t) \in l_2$$

$$s(t) = \sum_{n=1}^{\infty} \beta_n \sin(n\omega_0 t) \in l_2$$

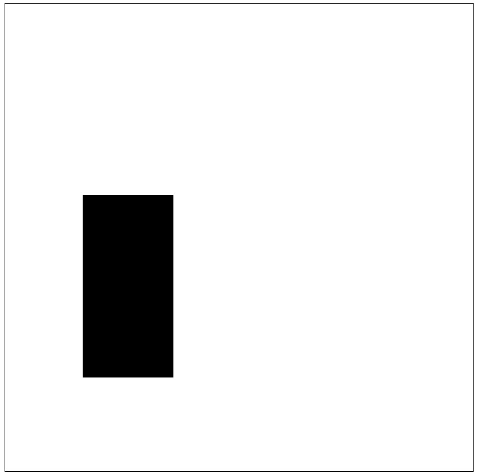


Vocal Tract modeled as a sequence of tubes.
(e.g. Stevens, 1998)

Vision Example

$$f : \mathbb{R}^2 \to [0, 1]$$

$$\mathcal{F} = \{f | f(x, y) = v(x - t, y - r)\}$$



Learning when data $\sim \mathcal{M} \subset \mathbb{R}^N$

- Clustering: $\mathcal{M} \to \{1, \ldots, k\}$

  connected components, min cut

- Classification: $\mathcal{M} \to \{-1, +1\}$

  P on $\mathcal{M} \times \{-1, +1\}$

- Dimensionality Reduction: $f : \mathcal{M} \to \mathbb{R}^n \quad n << N$

- $\mathcal{M}$ unknown: what can you learn about $\mathcal{M}$ from data?

  e.g. dimensionality, connected components

  holes, handles, homology

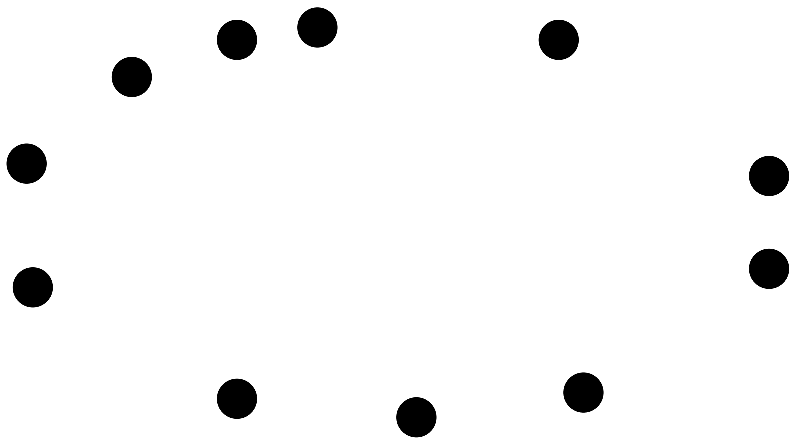
  curvature, geodesics

Dimensionality Reduction

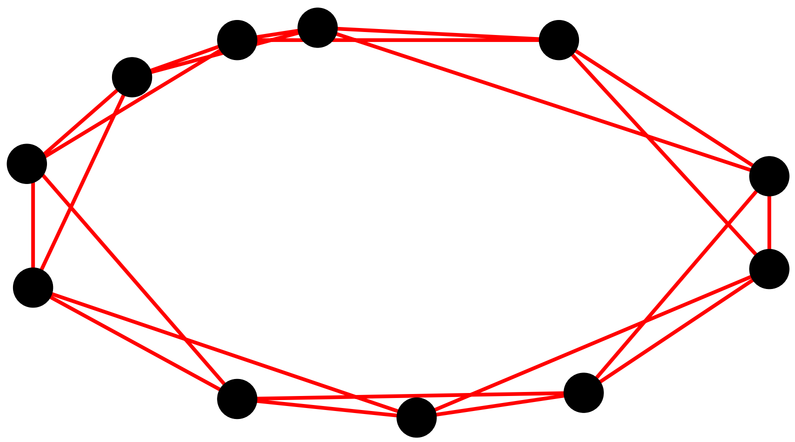
Given $x_1, \ldots, x_n \in \mathcal{M} \subset \mathbb{R}^N$,
Find $y_1, \ldots, y_n \in \mathbb{R}^d$ where $d << N$

- ISOMAP (Tenenbaum, et al, 00)
- LLE (Roweis, Saul, 00)
- Laplacian Eigenmaps (Belkin, Niyogi, 01)
- Hessian Eigenmaps (Donoho, Grimes, 02)
- Diffusion Maps (Coifman, Lafon, et al, 04)

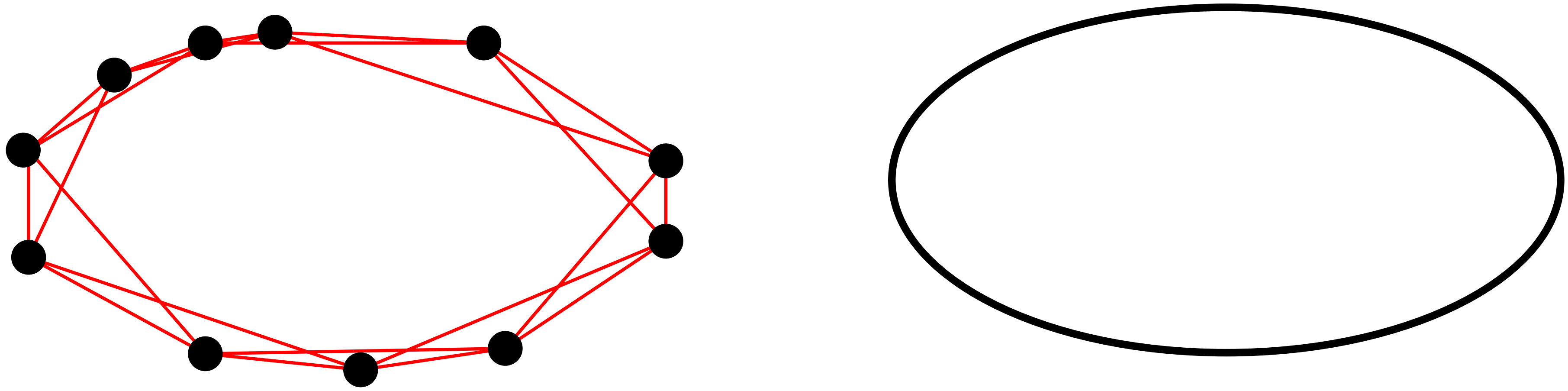
Algorithmic framework



Algorithmic framework



Algorithmic framework



Neighborhood graph common to all methods.

1. Construct Neighborhood Graph.
2. Find shortest path distances.

D_{ij} is $n \times n$

3. Embed using Multidimensional Scaling.

Multidimensional Scaling

Consider a positive definite matrix A.
Then A_{ij} corresponds to inner products.

$$A = \sum_{i=1}^{n} \lambda_i \phi_i \phi_i^T$$

Then for any $x \in \{1, \ldots, n\}$

$$\psi(x) = \left(\sqrt{\lambda_1}\phi_i(x), \ldots, \sqrt{\lambda_k}\phi_k(x)\right) \in \mathbb{R}^k$$

approximates inner products and therefore distances.
Therefore find A such that

$$A_{ii} + A_{jj} - 2A_{ij} \approx D_{ij}$$

Good Answer:

$$A = -\frac{1}{2}HDH \text{ where } H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$$

Step 1 [*Constructing the Graph*]

$$e_{ij} = 1 \Leftrightarrow \mathbf{x}_i \text{ “close to” } \mathbf{x}_j$$

1. ϵ-neighborhoods. [*parameter* $\epsilon \in \mathbb{R}$] Nodes i and j are connected by an edge if

$$||\mathbf{x}_i - \mathbf{x}_j||^2 < \epsilon$$

2. n nearest neighbors. [*parameter* $n \in \mathbb{N}$] Nodes i and j are connected by an edge if i is among n nearest neighbors of j or j is among n nearest neighbors of i.

Laplacian Eigenmaps

Step 2. [*Choosing the weights*].

1. Heat kernel. [*parameter* $t \in \mathbb{R}$]. If nodes i and j are connected, put

$$W_{ij} = e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{t}}$$

2. Simple-minded. [*No parameters*]. $W_{ij} = 1$ if and only if vertices i and j are connected by an edge.

Laplacian Eigenmaps

Step 3. [*Eigenmaps*] Compute eigenvalues and eigenvectors for the generalized eigenvector problem:

$$Lf = \lambda D f$$

D is diagonal matrix where

$$D_{ii} = \sum_j W_{ij}$$

$$L = D - W$$

Let $\mathbf{f}_0, \ldots, \mathbf{f}_{k-1}$ be eigenvectors.

Leave out the eigenvector $\mathbf{f}_0$ and use the next m lowest eigenvectors for embedding in an m-dimensional Euclidean space.

Justification

Find $y_1, \ldots, y_n \in R$

$$\min \sum_{i,j} (y_i - y_j)^2 W_{ij}$$

Tries to preserve locality

A Fundamental Identity

But

$$\frac{1}{2}\sum_{i,j}(y_i - y_j)^2 W_{ij} = \mathbf{y}^T L \mathbf{y}$$

$$\sum_{i,j}(y_i - y_j)^2 W_{ij} = \sum_{i,j}(y_i^2 + y_j^2 - 2y_i y_j) W_{ij}$$

$$= \sum_i y_i^2 D_{ii} + \sum_j y_j^2 D_{jj} - 2\sum_{i,j} y_i y_j W_{ij}$$

$$= 2\mathbf{y}^T L \mathbf{y}$$

$$\lambda = 0 \rightarrow \mathbf{y} = \mathbf{1}$$

$$\min_{\mathbf{y}^T \mathbf{1} = 0} \mathbf{y}^T L \mathbf{y}$$

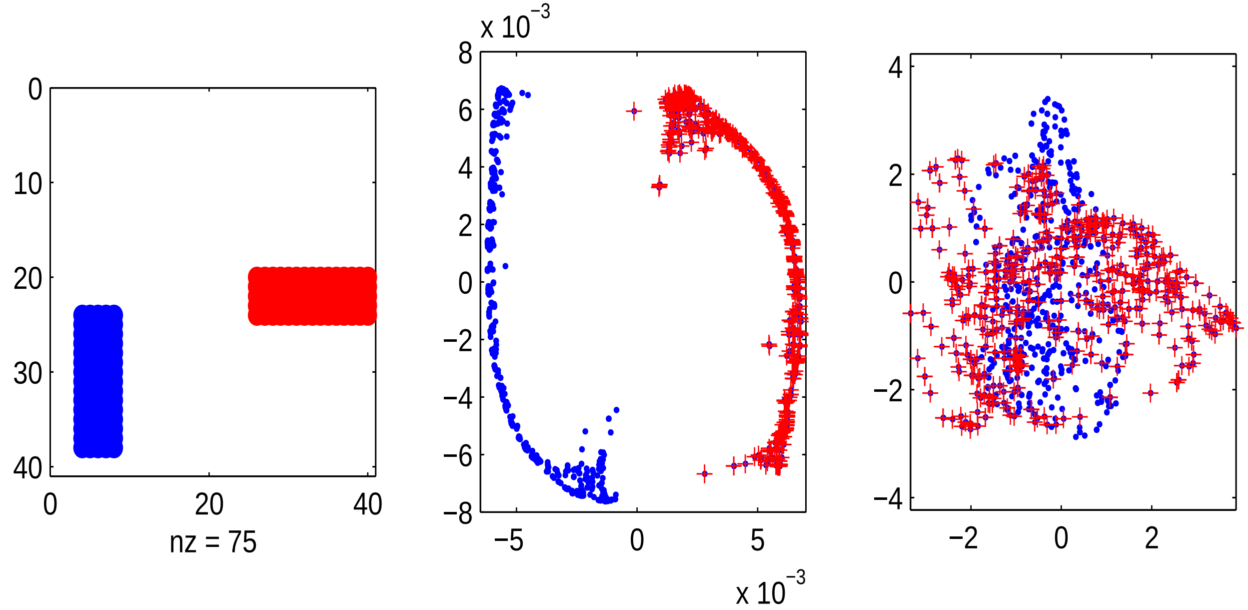
Let $Y = [\mathbf{y}_1 \mathbf{y}_2 \ldots \mathbf{y}_m]$

$$\sum_{i,j} ||Y_i - Y_j||^2 W_{ij} = \text{trace}(Y^T L Y)$$

subject to $Y^T Y = I$.

Use eigenvectors of L to embed.

PCA versus Laplacian Eigenmaps



smooth map $f : \mathcal{M} \to R$

$$\int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 \approx \sum_{i \sim j} W_{ij}(f_i - f_j)^2$$

Recall standard gradient in $\mathbb{R}^k$ of $f(z_1, \ldots, z_k)$

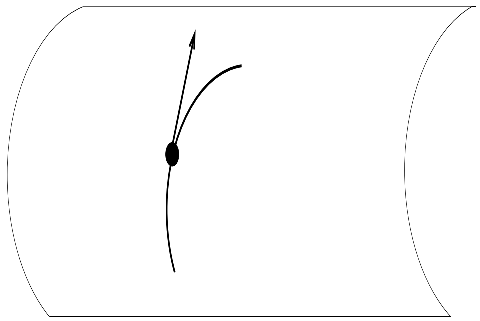
$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial z_1} \\ \frac{\partial f}{\partial z_2} \\ \cdot \\ \cdot \\ \frac{\partial f}{\partial z_k} \end{bmatrix}$$

Curves on Manifolds

Consider a curve on $\mathcal{M}$

$$c(t) \in \mathcal{M} \qquad t \in (-1,1) \qquad p = c(0);\; q = c(\tau)$$

$$f(c(t)) : (-1,1) \to \mathbb{R}$$



$$|f(0) - f(\tau)| \lesssim d_G(p,q) \|\nabla_M f(p)\|$$

A Basic Fact

$$\int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 = \int f \cdot \Delta_{\mathcal{M}} f$$

This is like

$$\sum_{i,j} W_{ij}(f_i - f_j)^2 = \mathbf{f}^T \mathbf{L} \mathbf{f}$$

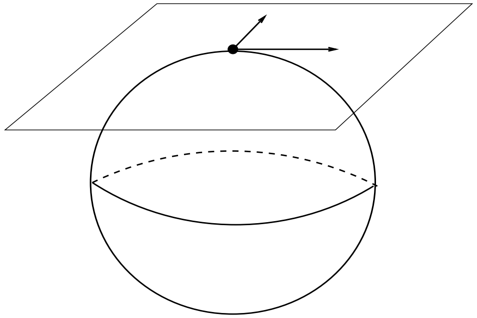
where

$\Delta_{\mathcal{M}} f$ is the manifold Laplacian

Recall ordinary Laplacian in $\mathbb{R}^k$
This maps

$$f(x_1, \ldots, x_k) \to \left(-\sum_{i=1}^{k} \frac{\partial^2 f}{\partial x_i^2} \right)$$

Manifold Laplacian is the same on the tangent space.



Properties of Laplacian

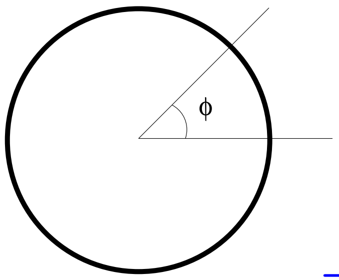
Eigensystem

$$\Delta_{\mathcal{M}} f = \lambda_i \phi_i$$

$\lambda_i \geq 0$ and $\lambda_i \to \infty$

$\{\phi_i\}$ form an orthonormal basis for $L^2(\mathcal{M})$

$$\int \|\nabla_{\mathcal{M}} \phi_i\|^2 = \lambda_i$$



$$-\frac{d^2u}{dt^2} = \lambda u \text{ where } u(0) = u(2\pi)$$

Eigenvalues are

$$\lambda_n = n^2$$

Eigenfunctions are

$$\sin(nt), \cos(nt)$$

From graphs to manifolds

$$f : \mathcal{M} \to \mathbb{R} \quad x \in \mathcal{M} \quad x_1, \ldots, x_n \in \mathcal{M}$$

Graph Laplacian:

$$L_n^t(f)(x) = f(x) \sum_j e^{-\frac{\|x - x_j\|^2}{t}} - \sum_j f(x_j) e^{-\frac{\|x - x_j\|^2}{t}}$$

Theorem 1 [pointwise convergence] $t_n = n^{-\frac{1}{k+2+\alpha}}$

$$\lim_{n \to \infty} \frac{(4\pi t_n)^{-\frac{k+2}{2}}}{n} L_n^{t_n} f(x) = \mathcal{L}_{\mathcal{M}} f(x)$$

Belkin 03, Belkin Niyogi 05

related work: Lafon Coifman 04, Hein et al 05

From graphs to manifolds

Theorem 2 [uniform convergence]

$$\lim_{n\to\infty} \sup_{x\in\mathcal{M},\, f\in\mathcal{B}} \left| \frac{(4\pi t_n)^{-\frac{k+2}{2}}}{n} L_n^{t_n} f(x) - \mathcal{L}_{\mathcal{M}} f(x) \right| = 0$$

Theorem 3 [convergence of eigenfunctions]

$$Eig[L_n^{t_n}] \to Eig[\mathcal{L}_{\mathcal{M}}]$$

Belkin Niyogi 05 [in preparation]